1 Wnioskowanie statystyczne — podstawowe pojęcia

1.1 Parametry rozkładu, próba losowa

We wnioskowaniu statystycznym próbujemy na podstawie losowej próbki z pewnej populacji wnioskować na temat całej populacji. Możemy na przykład zmierzyć wzrost 50 losowo wybranych studentów i na podstawie otrzymanych wyników wnioskować na temat wartości średniej czy wariancji tego wzrostu wśród wszystkich studentów.

Do modelowania takiej sytuacji używa się zazwyczaj następującego formalizmu. Zakładamy, że cecha, którą badamy ma pewien nieznany rozkład pochodzący ze znanej rodziny rozkładów $\mathcal{D} = \{\theta \in \Theta : \mathcal{P}_\theta\}$, gdzie $\mathcal{P}_\theta = (\Omega, \mathcal{F}, P_\theta)\}$ parametryzowanej przez parametr θ.

Chociaż nie znamy θ, a zatem także rozkładu $\mathcal{P}_\theta$, to jednak coś o nim wiemy. Znamy mianowicie wartości ciągu zmiennych losowych $X_1, \ldots, X_n$ o rozkładzie $\mathcal{P}_\theta$. Taki ciąg nazywamy *próbą losową*. Jeśli zmienne te są niezależne, to mówimy o próbie *prostej*.

Przykład: W przykładzie ze wzrostem studentów moglibyśmy założyć, że wzrost ma rozkład normalny o nieznanej wartości średniej μ i wariancji σ^2 (co oczywiście z wielu powodów nie ma szansy być prawdą, ale zapewne jest niezłym przybliżeniem). Moglibyśmy wtedy przyjąć $\Theta = (0, \infty) \times [0, \infty)$ i $\mathcal{D} = \{(\mu, \sigma^2) \in \Theta : \mathcal{N}(\mu, \sigma^2)\}$.

Wyniki pomiaru wzrostu 50 studentów możemy modelować zmiennymi losowymi $X_1, \ldots, X_{50}$. Wszystkie X_i mają ten sam rozkład $\mathcal{N}(\mu, \sigma^2)$, przy czym wartości μ ani σ^2 nie znamy. Ciąg $X_1, \ldots, X_{50}$ jest w tym przypadku próbą losową. O ile nie losujemy studentów "ze zwracaniem" próba ta w ogólnym przypadku nie jest prosta (dlaczego?), ale też nie popełnimy z reguły dużego błędu zakładając, że jest.

1.2 Podstawowe problemy

Naszym celem jest wnioskowanie na temat parametru θ na podstawie wartości zmiennych $X_1, \ldots, X_{50}$. Zajmiemy się m.in. następującymi trzema problemami:

Estymacja punktowa: Podać wartość parametru θ.

Estymacja przedziałowa: Podać (mały) przedział, który zawiera parametr θ.

Testowanie hipotez: Odpowiedzieć na pytanie w rodzaju: "Czy $\theta > a$?".

Oczywiście rozwiązanie pierwszego z tych problemów w prosty sposób prowadzi do rozwiązania pozostałych dwóch. Nie ma to jednak większego znaczenia, ponieważ żadnego z tych problemów w ogólnym przypadku rozwiązać się nie da. Można je jednak rozwiązywać "w przybliżeniu", tzn. opracować metody, które:

- podają "dobre" oszacowania wartości parametru θ,
- podają (mały) przedział, który z dużym prawdopodobieństwem zawiera parametr θ,
- pozwalają w "sensowny" sposób odpowiadać na pytania w rodzaju: "Czy $\theta > a$?".

(słowa "dobre" i "sensowny" zostaną sprecyzowane w dalszej części wykładu). Omówieniem takich właśnie metod zajmiemy się w kilku kolejnych wykładach.

1.3 Statystyki i estymatory

Zanim przejdziemy do estymacji punktowej, zdefiniujemy dwa fundamentalne pojęcia: statystykę i estymator.

Statystyką nazywamy dowolną funkcję $S(X_1, \ldots, X_n)$ próby losowej. Statystykami są na przykład $\min(X_1, \ldots, X_n)$, $\max(X_1, \ldots, X_n)$, $\frac{\sum_{i=1}^n X_i}{n}$, ale też $X_1 + 5X_3^2$ oraz $10003 + X_1 * X_2 - X_5$. Statystyki są oczywiście zmiennymi losowymi, określonymi na tej samej przestrzeni co $X_1, \ldots, X_n$.

Estymatorem parametru $\theta \in \Theta$ nazywamy dowolną statystykę przyjmującą wartości w Θ. Ta definicja nie niesie zbyt wiele treści. Intuicyjnie, estymator parametru θ to taka statystyka, której rzeczywiście używamy do szacowania wartości θ (wielu autorów używa tego sformułowania jako definicji estymatora, ale trudno nazwać je definicją).

Uwaga: Może się zdarzyć, że interesuje nas nie cały parametr θ, a jedynie niektóre jego składowe. Może nas, na przykład, interesować średni wzrost studentów, ale nie wariancja wzrostu. Pojęcia estymatora w oczywisty sposób przenosi się na tę sytuacją. Można też mowić o estymacji punktowej, przedziałowej i testowaniu hipotez.

2 Estymacja punktowa

W tym rozdziale postaramy się uściślić pojęcie "dobrego oszacowania", które pojawiło się we wcześniejszych rozważaniach. W ogólnym przypadku rozkład, z którego pochodzi próba losowa jest zależny od pewnego parametru θ, i ten właśnie parametr staramy się "dobrze oszacować". W tym rozdziale, dla uproszczenia, ograniczymy się do estymacji wartości oczekiwanej i wariancji, ale podobne rozważania można przeprowadzać także w innych przypadkach (patrz ćwiczenia).

2.1 Estymacja wartości oczekiwanej

W jaki sposób oszacować nieznaną wartość średnią populacji μ, jeśli mamy daną próbę losową $X_1, \ldots, X_n$. Naturalnym pomysłem jest użycie następującego estymatora

- $\hat{\mu}_1(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n X_i}{n}$,

ale każdy z poniższych pomysłów też wydaje się działać nienajgorzej

- $\hat{\mu}_2(X_1, \ldots, X_n) = X_1$,
- $\hat{\mu}_3(X_1, \ldots, X_n) = \frac{\sum_{i=1}^{n/2} X_i}{n/2}$,
- $\hat{\mu}_4(X_1, \ldots, X_n) = \sum_{i=1}^n \alpha_i X_i$, gdzie $\alpha_i > 0$ i $\sum_{i=1}^n \alpha_i = 1$,
- $\hat{\mu}_5(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n X_i}{n-1}$,
- $\hat{\mu}_6(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n X_i}{n} + 1$.

Każdy z estymatorów $\hat{\mu}_2$—$\hat{\mu}_6$ intuicyjnie wydaje się gorszy od $\hat{\mu}_1$, ale dla większości z nich trudno powiedzieć dlaczego.

Uwaga: Przyjęliśmy tu (i będziemy przyjmować w dalszej części tego wykładu) dość popularną w literaturze umowę polegającą na oznaczaniu estymatorów parametru θ symbolem $\hat{\theta}$, w tym przypadku estymatorów wartości oczekiwanej μ symbolem $\hat{\mu}$ z odpowiednim indeksem.

Uwaga: Każda z definicji $\hat{\mu}_1$ — $\hat{\mu}_6$ opisuje tak naprawdę cały ciąg estymatorów parametryzowaną rozmiarem próby n. Dla uproszczenia, nie będziemy tego faktu z reguły zaznaczać explicite. Warto o tym jednak pamiętać.

2.1.1 Estymatory nieobciążone

Obliczmy wartość oczekiwaną estymatora $\hat{\mu}_1$.

$$E\hat{\mu}_1(X_1, \ldots, X_n) = E\frac{\sum_{i=1}^n X_i}{n} = \frac{\sum_{i=1}^n EX_i}{n} = \frac{n\mu}{n} = \mu,$$

a zatem wartością oczekiwaną tego estymatora jest wartość estymowanej wielkości t.j. μ. Mówimy, że $\hat{\mu}_1$ jest *nieobciążonym* estymatorem μ.

W ogólnym przypadku mówimy, że $\hat{\theta}$ jest *nieobciążonym* (ang. unbiased) estymatorem θ, jeśli $E\hat{\theta} = \theta$. Różnicę $E\hat{\theta} - \theta$ nazywamy *obciążeniem* (ang. bias) estymatora, w przypadku estymatorów nieobciążonych jest ona równa 0.

Nieobciążoność wydaje się bardzo naturalną i pożądaną własnością estymatora. Zwróćmy uwagę, że estymatory $\hat{\mu}_5$ i $\hat{\mu}_6$ nie są nieobciążone i z tego właśnie powodu uważamy je za "gorsze" od $\hat{\mu}_1$.

W praktyce pojęcie nieobciążoności okazuje się jednak niejednokrotnie zbyt rygorystyczne co można zobaczyć na następującym przykładzie.

Zadanie Przeprowadzamy są badania rozkładu liczby telefonów na minutę w helpdesku. Zakładamy, że rozkład liczby telefonów na minutę jest rozkładem Poissona o nieznanym parametrze λ. Przypuśmy, że interesuje nas estymacja prawdopodobieństwa p tego, że w ciągu dwóch kolejnych minut nie odbierzemy żadnego telefonu. Pokaż, że dla próby jednoelementowej, jedyny estymator nieobciążony tego prawdopodobieństwa to $\hat{p}(X_1) = (-1)^{X_1}$.

Przykład ten pokazuje, że estymatory nieobciążone niekoniecznie są lepsze od obciążonych. W sytuacji z zadania dużo sensowniejszy jest na przykład estymator otrzymany metodą największej wiarygodności (omówioną w dalszej części wykładu) — $\hat{p}(X_1) = e^{-2X_1}$.

Dobrym osłabieniem pojęcia nieobciążoności jest tzw. asymptotyczna nieobciążoność. Mówimy, że estymator $\hat{\theta_n}(X_1, \ldots, X_n)$ (a właściwie ciąg estymatorów — zobacz uwaga w poprzednim podrozdziale) jest *asymptotycznie nieobciążony*, jeśli $\lim_{n\to\infty} E\hat{\theta}_n(X_1, \ldots, X_n) = \theta$.

Zwróćmy uwagę, że estymator $\hat{\mu}_5$ jest asymptotycznie nieobciążony chociaż nie jest nieobciążony.

2.1.2 Estymatory zgodne

W definicji asymptotycznej nieobciążoności dla estymatora $\hat{\theta}$ obserwowaliśmy zachowanie $E\hat{\theta}_n(X_1, \ldots, X_n)$ dla $n \to \infty$. Naturalne wydaje się jednak wymaganie, aby nie tylko $E\hat{\theta}_n(X_1, \ldots, X_n) \to \theta$, ale także $\hat{\theta}_n(X_1, \ldots, X_n) \to \theta$ dla $n \to \infty$.

Estymator $\hat{\theta}$ parametru θ nazywamy estymatorem *(słabo) zgodnym* jeśli dla dowolnego $\varepsilon > 0$

$$\lim_{n\to\infty} P(|\hat{\theta}_n(X_1, \ldots, X_n) - \theta| <= \varepsilon) = 1.$$

Uwaga: Definicja ta zapewne przypomina czytelnikowi (jak się za chwilę okaże słusznie) słabe prawo wielkich liczb. Można też wprowadzić mocną wersję tej definicji żądając aby

$$P(\lim_{n\to\infty} \hat{\theta}_n(X_1, \ldots, X_n) = \theta) = 1.$$

Sprawdźmy, czy nasze estymatory wartości oczekiwanej $\hat{\mu}_1$ — $\hat{\mu}_6$ są zgodne. Zacznijmy od $\hat{\mu}_1$. Musimy sprawdzić, czy dla dowolnego $\varepsilon > 0$ zachodzi

$$\lim_{n\to\infty} P(|\frac{\sum_{i=1}^n X_i}{n} - \mu| <= \varepsilon) = 1,$$

ale to jest dokładnie teza słabego prawa wielkich liczb (z mocnego prawa wielkich liczb wynika w analogiczny sposób, że $\hat{\mu}_1$ jest zgodny w mocnym sensie). Podobnie można pokazać, że zgodne są estymatory $\hat{\mu}_3$, $\hat{\mu}_4$ i $\hat{\mu}_5$. Nie jest zgodny estymator $\hat{\mu}_2$ i to jest jego główna wada — ponieważ estymator ten korzysta jedynie z wartości X_1 jakość estymacji nie poprawia się wraz ze wzrostem rozmiaru próby.

Zadanie Z naszych rozważań wynika, że estymator nieobciążony może nie być zgodny (estymator $\hat{\mu}_2$). Podaj przykład estymatora zgodnego, który nie jest (nawet asymptotycznie) nieobciążony.

2.1.3 Efektywność

Nasz dotychczasowe rozważania nie pozwalają porównać estymatorów $\hat{\mu}_1$, $\hat{\mu}_3$ i $\hat{\mu}_4$. Intuicyjnie wydaje się, że estymator $\hat{\mu}_1$ najefektywniej wykorzystuje wartości zmiennych $X_1, \ldots, X_n$ sumując wszystkie z jednakowymi współczynnikami. Jak za chwilę zobaczymy, ma dzięki temu najmniejszą wariancję.

Zacznijmy od obliczenia wariancji $\hat{\mu}_1$. Niech σ^2 będzie wariancją rozkładu, z którego pochodzi nasza próba $X_1, \ldots, X_n$. Wtedy

$$\mathrm{Var}(\hat{\mu}_1) = \mathrm{Var}\frac{\sum_{i=1}^n X_i}{n} = \frac{\mathrm{Var}\sum_{i=1}^n X_i}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Analogiczne obliczenia dla pozostałych dwóch estymatorów dają

$$\mathrm{Var}(\hat{\mu}_3) = \frac{\sigma^2}{n/2}$$

oraz

$$\mathrm{Var}(\hat{\mu}_4) = (\sum_{i=1}^n \alpha_i^2)\sigma^2.$$

Zadanie Pokaż, że wyrażenie opisujące wariancję estymatora $\hat{\mu}_4$ przyjmuje wartość najmniejszą, gdy wszystkie $\alpha_i = 1/n$, tzn. gdy $\hat{\mu}_4 = \hat{\mu}_1$.

Jeśli $\hat{\theta}_1$ i $\hat{\theta}_2$ są dwoma nieobciążonymi estymatorami pewnego parametru θ i $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$, to mówimy, że $\hat{\theta}_1$ jest *efektywniejszy* od $\hat{\theta}_2$. A zatem estymator $\hat{\mu}_1$ nie tylko intuicyjnie efektywniej wykorzystuje wartości zmiennych $X_1, \ldots, X_n$ ale też jest jest efektywniejszy od $\hat{\mu}_3$ i $\hat{\mu}_4$ w sensie powyższej definicji.

Warto w tym miejscu zwrócić uwagę, że porównywanie wariancji estymatorów o różnym obciążeniu niekoniecznie ma sens. Dlatego pojęcie efektywności definiujemy tylko dla estymatorów nieobciążonych. Czasem rozszerza się je na estymatory asymptotycznie nieobciążone.

W ogólnym przypadku wygodnie jest badać tzw. średni błąd kwadratowy estymatora (ang. mean square error, w skrócie MSE) zdefiniowany następująco

$$\mathrm{MSE}(\hat{\theta}) = E(\hat{\theta}(X_1, \ldots, X_n) - \theta)^2.$$

Oznaczając dla uproszczenia rachunków $\hat{\theta}(X_1, \ldots, X_n)$ przez $\hat{\theta}$ mamy

$$\mathrm{MSE}(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = E((\hat{\theta}-E(\hat{\theta}))+(E(\hat{\theta})-\theta))^2 = E(\hat{\theta}-E(\hat{\theta}))^2+2E(\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\theta)+E(E(\hat{\theta})-\theta)^2.$$

Pierwszy z wyrazów tej sumy jest wariancją $\hat{\theta}$, ostatni jest równy kwadratowi obciążenia $\hat{\theta}$, a środkowy, jak łatwo zauważyć, jest równy zeru. A zatem

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + (E\hat{\theta} - \theta)^2.$$

Podsumujmy nasze rozważania dotyczące estymatorów wartości oczekiwanej przypominając listę rozważanych estymatorów wraz z ich własnościami:

- $\hat{\mu}_1(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n X_i}{n}$, nieobciążony, zgodny, efektywny,
- $\hat{\mu}_2(X_1, \ldots, X_n) = X_1$, nieobciążony, nie zgodny,
- $\hat{\mu}_3(X_1, \ldots, X_n) = \frac{\sum_{i=1}^{n/2} X_i}{n/2}$, nieobciążony, zgodny, mniej efektywny niż $\hat{\mu}_1$,
- $\hat{\mu}_4(X_1, \ldots, X_n) = \sum_{i=1}^n \alpha_i X_i$, nieobciążony, zgodny, mniej efektywny niż $\hat{\mu}_1$ chyba, że wszystkie $\alpha_i = 1/n$, wtedy $\hat{\mu}_4 = \hat{\mu}_1$,

- $\hat{\mu}_5(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n X_i}{n-1}$, obciążony, ale asymptotycznie nieobciążony, zgodny,
- $\hat{\mu}_6(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n X_i}{n} + 1$, obciążony, nawet asymptotycznie, nie zgodny.

Uwaga: W statystyce definiuje się też formalnie tzw. *efektywność* nieobciążonego estymatora $\hat{\theta}$ parametru θ. Jest to iloraz $\frac{\mathrm{Var}(\hat{\theta}_0)}{\mathrm{Var}(\hat{\theta})}$, gdzie $\hat{\theta}_0$ jest najefektywniejszym (tzn. o najmniejszej wariancji) nieobciążonym estymatorem θ. Tak zdefiniowana efektywność jest liczbą z przedziału $[0, 1]$. Definicja ta wygląda na niezbyt przydatną, bo niby skąd możemy znać najefektywniejszy estymator θ? Okazuje się, że da się podać ograniczenie dolne na wariancję dowolnego nieobciążonego estymatora parametru θ. Jeśli nasz estymator ma taką właśnie wariancję, to wiemy, że jest najefektywniejszy. Osobom zainteresowanym tym wątkiem polecam znalazienie w dowolnym podręczniku statystyki (lub w Wikipedii) informacji na temat twierdzenia Rao-Cramera. Twierdzenie to niestety wykracza poza zakres tego wykładu.

2.2 Estymacja wariancji

Uzbrojeni w kryteria oceny estymatorów spróbujmy znaleźć dobry (tzn. nieobciążony, zgodny i efektywny) estymator dla wariancji. Naturalnym kandydatem wydaje się

$$\hat{\sigma^2}(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n},$$

gdzie μ jest wartością oczekiwaną. Oczywiście wzoru tego możemy użyć tylko w przypadku, gdy wartość oczekiwana μ jest z góry znana, a estymujemy jedynie wariancję. Sprawdźmy, czy $\hat{\sigma^2}$ jest w tym przypadku dobrym estymatorem wariancji. Mamy

$$E(\hat{\sigma^2}(X_1, \ldots, X_n)) = E\frac{\sum_{i=1}^n (X_i - \mu)^2}{n} = \frac{\sum_{i=1}^n E(X_i - \mu)^2}{n} = \frac{n\sigma^2}{n} = \sigma^2,$$

a zatem tak zdefiniowany estymator $\hat{\sigma^2}$ jest nieobciążony.

Zadanie Pokaż, że $\hat{\sigma^2}$ jest zgodnym estymatorem wariancji. W tym celu oblicz wariancję estymatora i użyj nierówności Czebyszewa.

Analiza efektywności tego estymatora w ogólnym przypadku nie jest łatwa. Można pokazać, że jeśli rodzina rozkładów z którą mamy do czynienia jest rodziną rozkładów normalnych, to powyższy estymator ma efektywność 1.

A co jeśli nie znamy wartości oczekiwanej? Naturalnym pomysłem jest użycie zamiast nieznanej wartości μ wartości estymatora $\hat{\mu}(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n X_i}{n}$, tj.

$$\hat{\sigma^2}(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n (X_i - \hat{\mu}(X_1, \ldots, X_n))^2}{n}.$$

Nie jest jednak jasne, czy tak zdefiniowany estymator jest nieobciążony. Aby to sprawdzić obliczmy jego wartość oczekiwaną (dla uproszczenia rachunków zamiast $\hat{\mu}(X_1, \ldots, X_n)$ będziemy pisać po prostu $\hat{\mu}$).

$$E\left(\frac{\sum_{i=1}^n (X_i - \hat{\mu})^2}{n}\right) = \frac{\sum_{i=1}^n E(X_i - \hat{\mu})^2}{n} = \frac{\sum_{i=1}^n E(X_i^2)}{n} - 2\frac{\sum_{i=1}^n E(X_i\hat{\mu})}{n} + \frac{\sum_{i=1}^n E(\hat{\mu}^2)}{n}.$$

Przede wszystkim zauważmy, że środkowy wyraz tej sumy można, korzystając z liniowości wartości oczekiwanej, przekształcić następująco

$$-2\frac{\sum_{i=1}^n E(X_i\hat{\mu})}{n} = -2E(\hat{\mu}^2),$$

a ostatni wyraz jest równy po prostu $E(\hat{\mu}^2)$. Po uproszczeniu dostajemy następujące wyrażenie na wartość oczekiwaną $\hat{\sigma^2}$

$$\frac{\sum_{i=1}^n E(X_i^2)}{n} - E(\hat{\mu}^2).$$

Do obliczenia wartości tego wyrażenia użyjemy znanej nam tożsamości

$$\mathrm{Var}X = E(X^2) - (EX)^2,$$

a raczej jej przekształconej wersji

$$E(X^2) = \mathrm{Var}X + (EX)^2.$$

Aby obliczyć wartość pierwszego wyrazu naszego wyrażenia wstawiamy do tożsamości X_i i dostajemy

$$\frac{\sum_{i=1}^n E(X_i^2)}{n} = \frac{\sum i = 1^n(\sigma^2 + \mu^2)}{n} = \sigma^2 + \mu^2.$$

Wstawiając do tej samej tożsamości $\hat{\mu}$ za X dostajemy wartość drugiego wyrazu

$$E(\hat{\mu}^2) = \mathrm{Var}(\hat{\mu}) + (E\hat{\mu})^2 = \frac{\sigma^2}{n} + \mu^2$$

(wartość $\mathrm{Var}(\hat{\mu})$ obliczyliśmy w poprzednim rozdziale).

Ostatecznie otrzymujemy

$$E(\hat{\sigma^2}) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2,$$

a zatem nasz estymator oblicza wariancję "prawie dobrze", wynik jest o czynnik $\frac{n-1}{n}$ za mały. Intuicyjnie przyczyną tego błędu estymacji jest to, że wartość $\hat{\mu}$ używana w tym estymatorze jest bliżej próby niż prawdziwa wartość μ.

Nasz obliczenia sugerują przy okazji sposób na otrzymanie nieobciążonego estymatora wariancji, wystarczy zastąpić n w mianowniku przez $n-1$:

$$\frac{\sum_{i=1}^n (X_i - \hat{\mu}(X_1, \ldots, X_n))^2}{n-1}.$$

Taki estymator jest nieobciążony, zgodny (udowodnienie tego faktu pozostawiamy czytelnikowi jako proste, ale żmudne, ćwiczenie) i w wielu sytuacjach "dość" efektywny (tzn. jego efektywność jest bliska 1, szczegółowe rozważania na ten temat wykraczają poza zakres tego wykładu).

2.3 Estymacja odchylenia standardowego

Estymacja odchylenia standardowego nie jest łatwa. Mogłoby się wydawać, że dobrym estymatorem odchylenia standardowego σ jest

$$\hat{\sigma} = \sqrt{\hat{\sigma^2}},$$

gdzie $\hat{\sigma^2}$ jest estymatorem wariancji (na przykład tym z poprzedniego podrozdziału).

Problem w tym, że nawet jeśli $\hat{\sigma^2}$ jest nieobciążonym estymatorem wariancji, to $\hat{\sigma}$ zdefiniowane jak powyżej nie musi być nieobciążonym estymatorem odchylenia standardowego. Można pokazać, że w takim przypadku (tj. gdy $\hat{\sigma^2}$ jest nieobciążony) zachodzi

$$E(\hat{\sigma}) \leq \sigma.$$

Zadanie Udowodnij to korzystając z tzw. nierówności Jensena, która mówi, że dla dowolnej funkcji f wypukłej do góry i dowolnej zmiennej losowej X, zachodzi

$$f(EX) \geq Ef(X).$$

(dowód tej nierówności w przypadku zmiennych dyskretnych w prosty sposób wynika z nierówności Jensena omawianej na analizie, to samo dotyczy przypadku zmiennych ciągłych choć dowód jest nieco bardziej skomplikowany).

Niestety, najczęściej $E(\hat{\sigma})$ jest mniejsze niż σ, tj. $\hat{\sigma}$ jest estymatorem obciążonym. Można temu zaradzić dodając do $\hat{\sigma}$ specjalne "poprawki", które pozwalają uzyskać nieobciążoność. W praktyce używa się tego estymatora bez zmian. Mimo, że jest on obciążony, obciążenie to nie jest duże, jest też asymptotycznie nieobciążony i zgodny.

2.4 Metoda największej wiarygodności

Jak dotąd omawialiśmy estymatory dla najbardziej naturalnych parametrów: wartości oczekiwanej, wariancji, odchylenia standardowego. Wzory, których używaliśmy, też były "naturalne" — staraliśmy się po prostu "naśladować" definicje odpowiednich pojęć. Można zapytać czy istnieją metody, które pozwalają w mniej lub bardziej automatyczny sposób konstruować dobre estymatory w ogólnym przypadku. Okazuje się, że tak, a najbardziej popularną metodą tego rodzaju jest tzw. metoda największej wiarygodności.

Idea metody największej wiarygodności jest bardzo prosta. Patrzymy na wartości zmiennych $X_1, \ldots, X_n$ i staramy się znaleźć taką wartość estymowanego parametru θ, dla której prawdopodobieństwo uzyskania tych właśnie wartości jest największe możliwe. Takie podejście działa w przypadku zmiennych o rozkładzie dyskretnym. Jeśli mamy do czynienia ze zmiennymi ciągłymi, to prawdopodobieństwo uzyskania dowolnych ustalonych wartości jest równe 0, dlatego metoda wymaga pewnych drobnych modyfikacji, ale idea pozostaje ta sama.

Podamy teraz formalną definicję funkcji wiarygodności oraz estymatora największej wiarygodności (ang. Maximum Likelihood Estimator (MLE)). Będziemy używać notacji z poprzednich podrozdziałów, przy czym w sytuacji, gdy $\mathcal{D}$ jest rodziną rozkładów ciągłych, będziemy dodatkowo używać symbolu f_θ na oznaczenie funkcji gęstości rozkładu odpowiadającego parametrowi θ.

Funkcją wiarygodności dla rodziny rozkładów parametryzowanej przez θ nazywamy:

- $L(\theta; x_1, \ldots, x_n) = P_\theta(X_1 = x_1) \cdot \ldots \cdot P_\theta(X_n = x_n)$ dla rozkładów dyskretnych,
- $L(\theta; x_1, \ldots, x_n) = f_\theta(x_1) \cdot \ldots \cdot f_\theta(x_n)$ dla rozkładów ciągłych.

Uwaga: W definicji tej zakładamy, tak jak się to z reguły robi, że mamy do czynienia z próbą prostą. Jeśli tak nie jest, należy użyć łącznego prawdopodobieństwa oraz łącznej funkcji gęstości.

Uwaga: Użycie symbolu θ w definicji funkcji L jest pewnego rodzaju nadużyciem notacyjnym, θ jest już przecież konkretną wartością parametru, którą próbujemy znaleźć. W praktyce użycie w tym miejscu tego samego symbolu nie prowadzi do niejasności i dla uproszczenia zapisu tak właśnie będziemy postępować w dalszej części tego podrozdziału.

Estymatorem największej wiarygodności (MLE) dla parametru θ nazywamy estymator $\hat{\theta}$, dla którego $\hat{\theta}(x_1, \ldots, x_n)$ jest równy takiej wartości parametru θ, dla której $L(\theta; x_1, \ldots, x_n)$ przyjmuje wartość największą.

Uwaga: W ogólnym przypadku może się zdarzyć, że wartość największa nie jest przyjmowana. Wtedy MLE nie jest określony. Może się też zdarzyć, że wartość największa jest przyjmowana w wielu punktach. Wtedy MLE jest określony niejednoznacznie.

Zobaczymy teraz dwa proste przykłady metody największej wiarygodności w akcji — jeden przykład dyskretny i jeden ciągły.

2.4.1 Metoda największej wiarygodności — przykład dyskretny

Rozważmy bardzo prosty przykład: $\mathcal{D}$ jest rodziną rozkładów Bernoulliego z nieznanym prawdopodobieństwem sukcesu s (wyjątkowo nie używamy zwyczajowego symbolu p w celu uniknięcia problemów notacyjnych), które chcemy estymować. W tym przypadku funkcja wiarygodności przyjmuje postać

$$L(s; x_1, \ldots, x_n) = P_s(X_1 = x_1) \cdot \ldots \cdot P_s(X_n = x_n) = s^k(1-s)^{n-k},$$

gdzie k jest liczbą jedynek w ciągu $x_1, \ldots, x_n$. Musimy teraz znaleźć wartość s dla której L przyjmuje wartość największą. Jeśli $k = 0$ to wartość największą L przyjmuje dla $s = 0$, podobnie gdy $k = n$, wartość największą L przyjmuje dla $s = 1$. W pozostałych przypadkach możemy obliczyć pochodną

$$L'(s; x_1, \ldots, x_n) = ks^{k-1}(1-s)^{n-k} - (n-k)s^k(1-s)^{n-k-1}.$$

Pochodna ta zeruje się dla $k(1-s) = (n-k)s$, czyli $s = k/n$ (uzasadnienie, że jest to globalne maksimum pominiemy). A zatem wartością estymatora MLE prawdopodobieństwa sukcesu s jest właśnie k/n. Warto zwrócić uwagę, że $k/n = \frac{\sum_{i=1}^n X_i}{n}$, a zatem w tym przypadku estymator MLE jest po prostu nieobciążonym estymatorem wartości oczekiwanej. Nie jest to specjalnie zaskakujące, skoro wartością oczekiwaną rozkładu Bernouliego z prawdopodobieństwem sukcesu s jest właśnie s. Należy to raczej traktować jako potwierdzenie skuteczności metody największej wiarygodności.

2.4.2 Metoda największej wiarygodności — przykład ciągły

W drugim przykładzie zajmiemy się sytuacją, w której $\mathcal{D}$ jest rodziną rozkładów normalnych $\mathcal{N}(\mu, \sigma^2)$ o nieznanej wartości oczekiwanej μ i wariancji σ^2. Znamy estymatory nieobciążone dla μ i σ^2. Interesującym pytaniem jest więc: czy metoda największej wiarygodności potrafi wygenerować te właśnie estymatory?

Funkcja wiarygodności wygląda w tym przypadku następująco

$$L(\mu, \sigma^2; x_1, \ldots, x_n) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}}.$$

Chcemy znaleźć wartości μ i σ^2 dla których L jest największe. Zauważmy, że niezależnie od wartości σ^2, najlepszą wartości μ jest taka wartość, dla której wyrażenie

$$\sum_{i=1}^n (x_i - \mu)^2$$

przyjmuje wartość najmniejszą. Różniczkując po μ dostajemy

$$-2\sum_{i=1}^n (x_i - \mu),$$

a przyrównując tę pochodną do zera otrzymujemy $\mu = \frac{\sum_{i=1}^n x_i}{n}$ i tę właśnie wartość μ zwróci estymator MLE wartości oczekiwanej. W dalszej części rachunków zakładamy, że μ ma wartość obliczoną przed chwilą.

Znajdziemy teraz wartość wariancji zwracaną przez estymator MLE. Dla uproszczenia obliczeń zamiast funkcji L będziemy rozpatrywać jej logarytm naturalny $\log L$ — możemy to zrobić, ponieważ L i $\log L$ przyjmują wartość największą dla tych samych wartości μ i σ^2. Jest to bardzo częsty zabieg w tego typu sytuacji, w nomenklaturze angielskiej ta zlogarytmowana funkcja nawet ma swoją nazwę: log-likelihood. W naszym przypadku dostajemy:

$$\log L(\mu, \sigma^2; x_1, \ldots, x_n) = -\frac{n}{2}\log\sqrt{2\pi} - n\log\sigma - \frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}.$$

Różniczkując to wyrażenie po σ dostajemy

$$-\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma^3},$$

a przyrównując pochodną do zera otrzymujemy

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n},$$

i tę właśnie wartość wariancji zwróci estymator (należy pamiętać, że μ jest tu średnią z próby). A zatem odpowiedź na pytanie zadane na początku tego podrozdziału brzmi: "prawie". Metoda największej wiarygodności daje znany nam nieobciążony estymator wartości oczekiwanej, ale w przypadku wariancji dostajemy estymator obciążony (ale asymptotycznie nieobciążony).

2.4.3 Własności estymatorów największej wiarygodności

Niech $\mathcal{D}$ będzie, jak zwykle, rodziną rozkładów prawdopodobieństwa parametryzowaną przez θ i niech $\gamma = f(\theta)$ dla pewnej bijekcji $f : \Theta \to \Gamma$. Wtedy możemy traktować $\mathcal{D}$ jako rodzinę rozkładów parametryzowaną przez γ.

Przykład: W przykładzie z rodziną rozkładów normalnych parametryzowanych przez wartość oczekiwaną μ i wariancję σ^2 moglibyśmy równie dobrze przyjąć, że naszymi parametrami są wartość oczekiwana i odchylenie standardowe. Wtedy $f(\mu, \sigma^2) = (\mu, \sigma)$.

Otóż w powyżej opisanej sytuacji, jeśli $\hat{\theta}$ jest estymatorem MLE dla parametru θ, to $f(\hat{\theta})$ jest estymatorem MLE dla $\gamma = f(\theta)$. Wynika to natychmiast z definicji estymatora MLE.

Przykład: Jeśli znamy estymator MLE dla wariancji, niech to będzie $\hat{\sigma^2}$, to estymatorem dla odchylenia standardowego jest $\sqrt{\hat{\sigma^2}}$.

Estymatory MLE, pomimo, że nie muszą być nieobciążone, mają wiele interesujących i pożądanych własności. Wiadomo na przykład, że przy dość słabych założeniach (które są jednak dość skomplikowane i nie możemy ich tu przytoczyć, założyć należy m.in. zgodność) estymatory MLE są m.in. asymptotycznie nieobciążone i asymptotycznie efektywne. Ta ostatnia własność oznacza, że ich wariancja dla $n \to \infty$ jest równa wariancji najefektywniejszego estymatora *nieobciążonego*.

Zadanie W rozdziale dotyczącym estymatorów nieobciążonych pokazaliśmy w przykładzie z helpdeskiem sytuację, w której jedyny estymator nieobciążony przyjmował wartości pozbawione sensu. Znajdź estymator MLE dla tego przykładu.

3 Estymacja przedziałowa

W tym rozdziale zajmiemy się tzw. estymacją przedziałową. Przy założeniach jak w poprzednim rozdziale, będziemy szukać nie pojedynczego estymatora $\hat{\theta}$ nieznanego parametru θ, ale pary statystyk θ_L i θ_R tak skonstruowanych, aby z dużym prawdopodobieństwem $\theta \in [\theta_L, \theta_R]$. Jeśli prawdopodobieństwo to jest $\geq 1 - \alpha$, to przedział $[\theta_L, \theta_R]$ nazywamy *przedziałem ufności* (dla θ) na poziomie ufności $1 - \alpha$.

Uwaga: Może zastanawiać użycie w powyższej definicji $1 - \alpha$, a nie po prostu α. Jest to standardowa notacja i, jak zobaczymy, prowadzi ona do nieco prostszych wzorów.

Uwaga: Warto w tym miejscu zwrócić uwagę na popularny błąd w rozumieniu powyższej definicji. A mianowicie, pomimo tego, że $P(\theta \in [\theta_L, \theta_R])$ to po obliczeniu $[\theta_L(x_1, \ldots, x_n), \ldots, \theta_R(x_1, \ldots, x_n)]$

dla konkretnych wartości $x_1, \ldots, x_n$ nie możemy twierdzić, że przedział ten zawiera z prawdopodobieństwem $1-\alpha$ prawdziwą wartość θ. Wartość ta bowiem albo należy do naszego przedziału albo nie, nie ma tu żadnej losowości. Prawdopodobieństwo sukcesu $1-\alpha$ dotyczy sytuacji *przed* wylosowaniem próby i wykonaniem obliczeń, dla frakcji $1-\alpha$ wszystkich możliwych prób odniesiemy sukces.

Podana przez nas definicja przedziału ufności w żaden sposób nie ogranicza wartośc θ_L i θ_R. Jasne jest jednak, że chcielibyśmy, aby przedział $[\theta_L, \theta_R]$ był możliwie krótki. Czasem interesuje nas także znalezienie jak najlepszego (trudno tu mówić o dlugości) przedziału postaci $[\theta_L, infty)$ lub $(-\infty, \theta_R]$. Są to tzw. *jednostronne* przedziały ufności.

Omówimy teraz estymację przedziałową wartości oczekiwanej i wariancji w 4 typowych sytuacjach. Techniki, których użyjemy są jednak dość ogólne i przenoszą się na dużą część sytuacji spotykanych w praktyce.

3.1 Sytuacja 1 — Estymacja wartości oczekiwanej rozkładu normalnego o znanej wariancji

Sytuacja ta jest do pewnego stopnia sztuczna, rzadko bowiem chcemy estymować wartość oczekiwaną znając wariancję. Jest tak na przykład wtedy gdy dokonujemy jakiegoś pomiaru, którego dokładność z góry znamy. Możemy na przykład ważyć pewien przedmiot za pomocą wagi, której błąd ma rozkład normalny o znanej wariancji (co jest dość mocnym, ale nie absurdalnym założeniem :)) Przede wszystkim jednak, omawiany przypadek ma charakter teoretyczny — rozważania są w nim szczególnie proste i stanowią dobry model sposobu postępowania w zbliżonych, ale nie co trudniejszych sytuacjach.

Niech $X_1, \ldots, X_n$ będzie próbą losową. Wiemy, że $\hat{\mu}(X_1, \ldots, X_n) = \frac{\sum_{i=1}^n X_i}{n}$ jest nieobciążonym estymatorem μ i naturalne byłoby skonstruowanie przedziału ufności wokół $\hat{\mu}$. Jak długi powinien być ten przedział?

Ponieważ wszystkie X_i mają rozkład normalny $\mathcal{N}(\mu, \sigma^2)$, to $\hat{\mu}$ ma rozkład $\mathcal{N}(\mu, \frac{\sigma^2}{n})$. W takim razie zmienna $Z = \frac{(\hat{\mu}-\mu)\sqrt{n}}{\sigma}$ ma rozkład $\mathcal{N}(0,1)$.

Niech Φ będzie dystrybuantą standardowego rozkładu normalnego $\mathcal{N}(0,1)$. Wtedy

$$P\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) \leq Z \leq \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha.$$

Korzystając z symetrii rozkładu normalnego mamy $\Phi^{-1}\left(\frac{\alpha}{2}\right) = -\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$, a wstawiając definicję Z dostajemy

$$P\left(-\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \leq \frac{(\hat{\mu}-\mu)\sqrt{n}}{\sigma} \leq \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha.$$

Przekształcamy nierówności, aby uzyskać przedział ufności dla μ

$$P\left(-\frac{\sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{\sqrt{n}} + \mu \leq \hat{\mu} \leq \frac{\sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{\sqrt{n}} + \mu\right) = 1-\alpha,$$

i ostatecznie

$$P\left(\hat{\mu} - \frac{\sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{\sqrt{n}} \leq \mu \leq \hat{\mu} + \frac{\sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{\sqrt{n}}\right) = 1-\alpha,$$

a zatem szukanym przedziałem ufności na poziomie $1-\alpha$ jest $\left[\hat{\mu} - \frac{\sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{\sqrt{n}}, \hat{\mu} + \frac{\sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{\sqrt{n}}\right]$. Łatwo sprawdzić, że jednostronnymi przedziałami ufności na poziomie ufności odpowiednio $1-\alpha$ są $\left(-\infty, \hat{\mu} + \frac{\sigma\Phi^{-1}(1-\alpha)}{\sqrt{n}}\right]$ oraz $\left[\hat{\mu} - \frac{\sigma\Phi^{-1}(1-\alpha)}{\sqrt{n}}, \infty\right)$. W kolejnych przykładach będziemy pomijać wzory na przedziały jednostronne. Czytelnik może je jednak łatwo wyprowadzić.

Uwaga: Nie jest być może do końca jasne po co normalizujemy zmienną $\hat{\mu}$ definiując zmienną Z. Korzystając z dowolnego programu statystycznego (np. R) można bowiem odczytać wartości odwrotności dystrybuanty (czyli po prostu kwantyle) rozkładu $\mathcal{N}(\mu, \sigma^2)$ dla dowolnych μ i σ^2. Normalizacja ma w dużej mierze charakter historyczny — w tablicach matematycznych, których kiedyś używano w tego typu obliczeniach, znajdują się jedynie kwantyle standardowego rozkładu normalnego.

3.2 Sytuacja 2 — Estymacja wartości oczekiwanej rozkładu normalnego o nieznanej wariancji

Sytuacja, którą teraz omówimy należy do najważniejszych i najbardziej typowych w praktyce.

Nie możemy powtórzyć bez zmian rozważań z poprzedniego podrozdziału, ponieważ nie znamy wartości σ występującej w definicji zmiennej Z. Narzuca się użycie zamiast σ wartości $\sqrt{\hat{\sigma^2}}$, gdzie $\hat{\sigma^2}$ jest nieobciążonym estymatorem wariancji zdefiniowanym w poprzednim rozdziale, a zatem

$$Z = \frac{(\hat{\mu} - \mu)\sqrt{n}}{\hat{\sigma^2}}.$$

Nie jest jednak jasne, czy tak zdefiniowana zmienna Z ma tak jak poprzednio rozkład normalny. Okazuje się, że nie. Z ma rozkład zwany *rozkładem t Studenta o $\nu = n - 1$ stopniach swobody* (rozkład ma stopnie swobody, a nie Student). Rozkład ten jest bardzo podobny do normalnego (i dla $\nu \to \infty$ zbiega do rozkładu normalnego), ale jego funkcja gęstości ma nieco grubsze ogony i niższe centrum (dlaczego?). Nie będziemy podawać wzoru na gęstość tego rozkładu, nie jest on prosty. Warto zaznaczyć, że rozkład Studenta jest jednym z najważniejszych rozkładów w statystyce i każdy sensowny pakiet statystyczny pozwala obliczać jego kwantyle, próbkować z niego itp.

Uwaga: Statystykę Z często oznacza się symbolem t, ze względu na jej rozkład.

Dalsza część rozważań w tym podrozdziale jest analogiczna do podrozdziału poprzedniego. Jeśli przez F_{n-1} oznaczymy dystrybuantę rozkładu Studenta o $n-1$ stopniach swobody, to dostajemy następujący przedział ufności na poziomie ufności $1 - \alpha$

$$\left[\hat{\mu} - \frac{\sigma F_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}, \hat{\mu} + \frac{\sigma F_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}\right]$$

3.3 Sytuacja 3 — Estymacja wartości oczekiwanej dowolnego rozkładu, duży rozmiar próby

Tym razem rodzina rozkładów z którą mamy do czynienie niekoniecznie jest rodziną rozkładów normalnych, a wariancja, tak jak w poprzedniej sytuacji, nie jest znana. Zakładamy jednak, że mamy do czynienia z dużą próbą, np. $n \geq 100$.

W tej sytuacji, ze względu na duży rozmiar próby, możemy przyjąć, że:

- $\hat{\mu}(X_1, \ldots, X_n)$ ma rozkład normalny (na mocy Centralnego Twierdzenia Granicznego zmienna ta ma rozkład bardzo bliski normalnemu),
- $\sqrt{\hat{\sigma^2}} = \sigma$.

Jeśli przyjmiemy powyższe założenia, to zmienna $Z = \frac{(\hat{\mu}-\mu)\sqrt{n}}{\sqrt{\hat{\sigma^2}}}$ ma rozkład $\mathcal{N}(0, 1)$ i możemy skorzystać z rozumowania z sytuacji 1. Dostajemy w ten sposób następujący przedział ufności na poziomie ufności $1 - \alpha$

$$\left[\hat{\mu} - \frac{\sqrt{\hat{\sigma^2}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}, \hat{\mu} + \frac{\sqrt{\hat{\sigma^2}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}\right]$$

3.4 Sytuacja 4 — Estymacja wariancji rozkładu normalnego, wartość oczekiwana nieznana

Zanim przejdziemy do konstruowania przedziału ufności dla wariancji zdefiniujemy kolejny rozkład prawdopodobieństwa. *Rozkładem χ^2_n (chi-kwadrat o n stopniach swobody)* nazywamy rozkład sumy $X_1^2+\ldots+X_n^2$, gdzie X_i niezależne o rozkładzie $\mathcal{N}(0,1)$. Podobnie jak w przypadku rozkładu Studenta pominiemy szczegółowe wzory opisujące ten rozkład. Będziemy natomiast korzystać z następującego twierdzenia (które podajemy bez dowodu)

Twierdzenie Jeśli $X_1,\ldots,X_n$ są niezależnymi zmiennymi o rozkładzie $\mathcal{N}(\mu,1)$ (μ jest tu dowolne), a $\bar{X}=\frac{\sum_{i=1}^n X_i}{n}$ jest ich średnią arytmetyczną, to zmienna $\sum_{i=1}^n (X_i-\bar{X})^2$ ma rozkład χ^2_{n-1}.

Popatrzmy teraz na zdefiniowany w rozdziale o estymacji punktowej nieobciążony estymator wariancji

$$\hat{\sigma^2}(X_1,\ldots,X_n)=\frac{\sum_{i=1}^n (X_i-\hat{\mu}(X_1,\ldots,X_n))}{n-1}.$$

Łatwo zauważyć, że

$$Z=\frac{\hat{\sigma^2}(X_1,\ldots,X_n)(n-1)}{\sigma^2}=\sum_{i=1}^n\left(\frac{X_i}{\sigma}-\frac{\hat{\mu}(X_1,\ldots,X_n)}{\sigma}\right)^2$$

ma na mocy twierdzenia rozkład χ^2_{n-1} (zmienne $\frac{X_i}{\sigma}$ mają rozkład normalny o wariancji 1).

Dalej postępujemy analogicznie jak poprzednio, ale ponieważ obliczenia mają tym razem nieco inną postać to podamy je w całości. Niech G_{n-1} będzie dystrybuantą rozkładu χ^2_{n-1}. Wtedy, podobnie jak poprzednio, mamy

$$P\left(G_{n-1}^{-1}\left(\frac{\alpha}{2}\right)\le Z\le G_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)\right)=1-\alpha.$$

Wstawiając definicję zmiennej Z dostajemy

$$P\left(G_{n-1}^{-1}\left(\frac{\alpha}{2}\right)\le\frac{(n-1)\hat{\sigma^2}}{\sigma^2}\le G_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)\right)=1-\alpha,$$

a przekształcając

$$P\left(\frac{G_{n-1}^{-1}\left(\frac{\alpha}{2}\right)}{(n-1)\hat{\sigma^2}}\le\frac{1}{\sigma^2}\le\frac{G_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)}{(n-1)\hat{\sigma^2}}\right)=1-\alpha,$$

i wreszcie

$$P\left(\frac{(n-1)\hat{\sigma^2}}{G_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)}\le\sigma^2\le\frac{(n-1)\hat{\sigma^2}}{G_{n-1}^{-1}\left(\frac{\alpha}{2}\right)}\right)=1-\alpha,$$

co daje nam szukany przedział ufności na poziomie ufności $1-\alpha$

$$\left[\frac{(n-1)\hat{\sigma^2}}{G_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)},\frac{(n-1)\hat{\sigma^2}}{G_{n-1}^{-1}\left(\frac{\alpha}{2}\right)}\right].$$

Warto zwrócić uwagę na to, że rozkład χ^2_{n-1} nie jest symetryczny, w związku z czym potrzebujemy dwóch różnych kwantyli we wzorach na przedział ufności.

4 Testowanie hipotez