Komputerowa analiza danych doświadczalnych

Wykład 9
6.05.2022

dr inż. Łukasz Graczykowski
lukasz.graczykowski@pw.edu.pl

Semestr letni 2021/2022

Metoda największej wiarygodności

Próby z rozkładów cząstkowych

Rozkłady t-Studenta i χ^2

Testy statystyczne

Jednoczesna estymacja kilku parametrów

Jednoczesna estymacja kilku parametrów

- Jak już pamiętamy, w najogólniejszym przypadku mamy pewną ilość p parametrów $\boldsymbol{\lambda}=(\lambda_{1,}\lambda_{2,\ldots},\lambda_p)$, które chcemy estymować. Wtedy funkcja wiarygodności i równania wiarygodności:

$$L=\prod_{j=1}^{N} f\left(\boldsymbol{X}^{(j)};\boldsymbol{\lambda}\right) \qquad l=\ln L=\sum_{j=1}^{N}\ln f\left(\boldsymbol{X}^{(j)};\boldsymbol{\lambda}\right) \qquad \frac{\partial l}{\partial\lambda_i}=0,\quad i=1,2,\ldots,p$$

- Niestety, w przypadku wielu parametrów wszystko nam się komplikuje dodatkowo z uwagi na możliwe **korelacje pomiędzy parametrami**

- Analogicznie jak dla 1 parametru, rozwijamy funkcję wiarygodności:

$$L\left(X^{(1)},X^{(2)},\ldots,X^{(j)};\boldsymbol{\lambda}\right)=\sum_{j=1}^{N}\ln f\left(X^{(j)};\boldsymbol{\lambda}\right)$$

na szereg Taylora wokół punktu:

$$\tilde{\boldsymbol{\lambda}}=(\tilde{\lambda}_{1,}\tilde{\lambda}_{2,\ldots},\tilde{\lambda}_p)$$

- Korzystając ze znikania 1-szych pochodnych dostajemy wtedy:

$$-\left(l(\boldsymbol{\lambda})-l(\tilde{\boldsymbol{\lambda}})\right)=1/2(\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}})^T A(\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}})+\ldots$$

- Gdzie macierz A:

$$-A=\begin{pmatrix} \frac{\partial^2 l}{\partial\lambda_1^2} & \frac{\partial^2 l}{\partial\lambda_1\partial\lambda_2} & \cdots & \frac{\partial^2 l}{\partial\lambda_1\partial\lambda_p} \\ \frac{\partial^2 l}{\partial\lambda_2\partial\lambda_1} & \frac{\partial^2 l}{\partial\lambda_2^2} & \cdots & \frac{\partial^2 l}{\partial\lambda_2\partial\lambda_p} \\ \frac{\partial^2 l}{\partial\lambda_p\partial\lambda_1} & \frac{\partial^2 l}{\partial\lambda_p\partial\lambda_2} & \cdots & \frac{\partial^2 l}{\partial\lambda_p^2} \end{pmatrix}$$

Jednoczesna estymacja kilku parametrów

- W granicy $N \rightarrow \infty$ można zastąpić macierz A odpowiednimi wartościami oczekiwanymi (niezależnymi od próby) $B = E(A)$
  - de facto zamieniamy $\lambda_{1,} \lambda_{2,} \ldots, \lambda_p$ na $\tilde{\lambda}_{1,} \tilde{\lambda}_{2,} \ldots, \tilde{\lambda}_p$
- Jeśli zaniedbamy wyrazy wyższych rzędów w rozwinięciu na szereg Taylora, dostaniemy wówczas funkcję wiarygodności postaci:

$$L = k \exp\left(-1/2(\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}})^T B(\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}})\right) \qquad C = B^{-1}$$

- Analogicznie jak dla jednego parametru, mamy tutaj p-wymiarowy rozkład normalny z macierzą kowariancji (dla estymatorów)
- Wariancje estymatorów największej wiarygodności dane są przez elementy diagonalne macierzy C: $\sigma^2(\tilde{\lambda}_i) = c_{ii}$
- Elementy pozadiagonalne są kowariancjami poszczególnych par estymatorów: $cov(\tilde{\lambda}_j, \tilde{\lambda}_k) = c_{jk}$ $\quad \tilde{\lambda}_{1,} \tilde{\lambda}_{2,} \ldots, \tilde{\lambda}_p$
- Możemy zdefiniować **współczynnik korelacji między estymatorami**: $\rho(\tilde{\lambda}_j, \tilde{\lambda}_k) = \frac{cov(\tilde{\lambda}_j, \tilde{\lambda}_k)}{\sigma(\tilde{\lambda}_j)\sigma(\tilde{\lambda}_k)}$

Jednoczesna estymacja kilku parametrów

- Analogicznie jak w przypadku z jednym parametrem, niepewnością estymacji (typu A) będzie pierwiastek kwadratowy z wariancji estymatora:

$$u(\tilde{\lambda}_i)=\sigma(\tilde{\lambda}_i)=\sqrt{c_{ii}}$$

- W przypadku 1D stwierdziliśmy, że za pomocą estymatora i jego niepewności możemy zdefiniować przedział obejmujący wartość prawdziwą estymowanego parametru z prawdopodobieństwem 68,3%. W przypadku wielowymiarowym przedział ten będzie określony również przez pełną macierz kowariancji (zależności między parametrami)
- Innymi słowy: obszar taki sprowadza się do elipsy kowariancji, analogicznej jak dla zmiennych losowych – czyli mamy **elipsy kowariancji dla estymowanych parametrów naszego rozkładu**
- Równanie **elipsoidy kowariancji** (p-wymiarowa przestrzeń):

$$g(\lambda)=1=2\{l(\lambda)-l(\tilde{\lambda})\}=(\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}})^T B(\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}})$$

- Dla $g(\lambda)=1$ mamy obszar ufności z prawdopodobieństwem 68,3%

Jednoczesna estymacja kilku parametrów - przykład

Jednoczesna estymacja kilku parametrów

- **Przykład:** badamy zasięg cząstek α w materii powstałych w wyniku rozpadu promieniotwórczego. Zasięg podlega rozkładowi normalnemu wokół wartości średniej → Innymi słowy, w wyniku eksperymentu mamy próbę losową o liczebności N z rozkładu normalnego
- Funkcja wiarygodności w tym przypadku:

$$L=\prod_{j=1}^{N}\frac{1}{\lambda_2\sqrt{2\pi}}\exp\frac{-\left(X^{(j)}-\lambda_1\right)^2}{2\lambda_2^2}$$

$$l=\ln L=-\frac{1}{2}\sum_{j=1}^{N}\frac{\left(X^{(j)}-\lambda_1\right)^2}{2\lambda_2^2}-N\ln\lambda_2-const$$

- Układ równań wiarygodności:

$$\frac{\partial l}{\partial\lambda_1}=\sum_{j=1}^{N}\frac{X^{(j)}-\lambda_1}{2\lambda_2^2}=0$$

$$\frac{\partial l}{\partial\lambda_2}=\frac{1}{\lambda_2^3}\sum_{j=1}^{N}\left(x^{(j)}-\lambda_1\right)^2-\frac{N}{\lambda_2}=0$$

- Rozwiązanie:

$$\tilde{\lambda}_1=\frac{1}{N}\sum_{j=1}^{N}X^{(j)}$$

$$\tilde{\lambda}_2=\sqrt{\frac{\sum_{j=1}^{N}\left(X^{(j)}-\lambda_1\right)^2}{N}}$$

Jednoczesna estymacja kilku parametrów

- Wyznaczamy macierz A, a następnie B i macierz C poprzez policzenie drugich pochodnych i w granicy $N \to \infty$ zastąpienie ich wartościami oczekiwanymi:

$$\frac{\partial^2 l}{\partial \lambda_1^2} = -\frac{N}{\lambda_2^2} \qquad \frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_2} = -\frac{2\sum X^{(j)} - \lambda_1}{\lambda_2^3} \qquad \frac{\partial^2 l}{\partial \lambda_2^2} = -\frac{3\sum \left(X^{(j)} - \lambda_1\right)^2}{\lambda_2^4} + \frac{N}{\lambda_2^2}$$

$$B = \begin{pmatrix} N/\tilde{\lambda}_2^2 & 0 \\ 0 & 2N/\tilde{\lambda}_2^2 \end{pmatrix} \qquad C = B^{-1} = \begin{pmatrix} \tilde{\lambda}_2^2/N & 0 \\ 0 & \tilde{\lambda}_2^2/2N \end{pmatrix}$$

- Elementy diagonalne macierzy C to odchylenia standardowe estymowanych parametrów
- Elementy pozadiagonalne wynoszą 0, zatem brak jest korelacji między estymowanymi parametrami

$$-\left(l(\boldsymbol{\lambda}) - l(\tilde{\boldsymbol{\lambda}})\right) = 1/2(\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}})^T A (\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}) + \ldots$$

$$L = k \exp\left(-1/2(\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}})^T B (\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}})\right) \qquad C = B^{-1}$$

$$B = E(A), \text{ dla } N \to \infty, \boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}$$

$$-A = \begin{pmatrix} \frac{\partial^2 l}{\partial \lambda_1^2} & \frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_2} & \cdots & \frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_p} \\ \frac{\partial^2 l}{\partial \lambda_2 \partial \lambda_1} & \frac{\partial^2 l}{\partial \lambda_2^2} & \cdots & \frac{\partial^2 l}{\partial \lambda_2 \partial \lambda_p} \\ \frac{\partial^2 l}{\partial \lambda_p \partial \lambda_1} & \frac{\partial^2 l}{\partial \lambda_p \partial \lambda_2} & \cdots & \frac{\partial^2 l}{\partial \lambda_p^2} \end{pmatrix}$$

7.1 The Fit Method

The Fit method is implemented in ROOT for the histogram classes `TH1`, the sparce histogram classes, `THnSparse`, the graph classes, `TGraph`, `TGraph2D` and `TMultiGraph` for fitting a collection of Graphs with the same function.

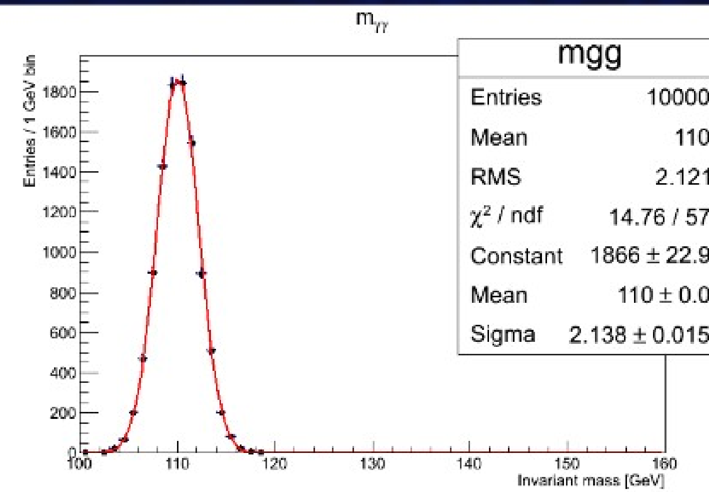
7.1.1 The TH1::Fit Method

To fit a histogram programmatically, you can use the `TH1::Fit` method. Here is the signatures of `TH1::Fit` and an explanation of the parameters:

```
TFitResultPtr Fit(TF1 *function, Option_t *option, Option_t *goption,
          Axis_t xxmin, Axis_t  xxmax)
```

- `function` a pointer to the fitted function (the fit model) object. One can also use the function name. This name may be one of ROOT pre-defined function names or a user-defined function. See the next paragraph for the list of pre-defined functions.
- `*option:` The second parameter is the fitting option. Here is the list of fitting options:
  - " `W` " Set all weights to 1 for non empty bins; ignore error bars
  - " `WW` " Set all weights to 1 including empty bins; ignore error bars
  - " `I` " Use integral of function in bin instead of value at bin center
  - " `L` " Use log likelihood method (default is chi-square method). To be used when the histogram represents counts
  - " `WL` " Weighted log likelihood method. To be used when the histogram has been filled with weights different than 1.
  - " `P` " Use Pearson chi-square method, using expected errors instead of the observed one given by `TH1::GetBinError` (default case). The expected error is instead estimated from the the square-root of the bin function value.

Example: Likelihood fit



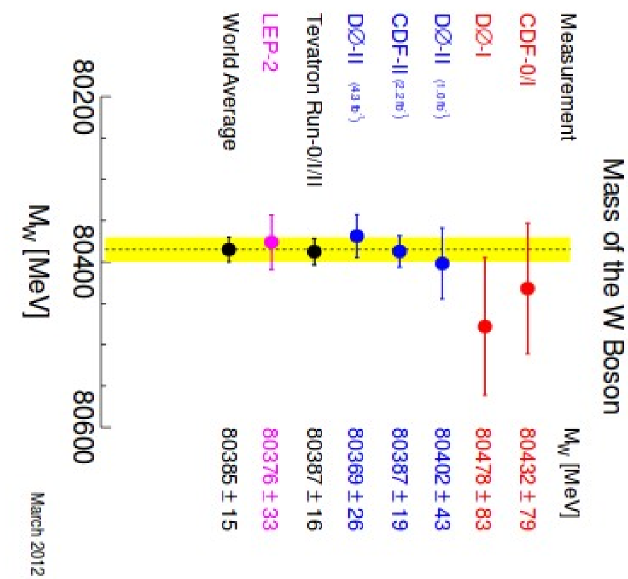
- ***Use the pdf of observables and maximize the product of likelihoods.***
  - In practice: minimize the -2*logarithm of the likelihoods
- **Binned data**
  - Use Poisson pdf for bin content
  - **Example:** data normally distributed

$$\mathcal{L} = \prod_{i=\text{bins}} \text{Poisson}\left[N_i^{\text{obs}}, A\exp\left(-\frac{1}{2}\frac{(x_i-\bar{x})^2}{\sigma^2}\right)\right]$$

- **Unbinned data**
  - **Example:** normally distributed points about α

$$\mathcal{L} = \prod_{i=\text{measurements}} \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{1}{2}\frac{(x_i-\alpha)^2}{\sigma_i^2}\right)$$

$$-2\ln\mathcal{L} = \sum_{i=\text{measurements}} \left[\frac{(x_i-\alpha)^2}{\sigma_i^2} + 2\ln\left(\sqrt{2\pi}\sigma_i\right)\right]$$



  - It looks familiar, doesn't it?
- **1σ uncertainty**: increase of $-2\ln\mathcal{L}$ by 1

Simple fits

```
myh2->Fit("gaus","","",-2.,2.);
```

Function name | Fit options | Plotting options | Fit range

```
myh2->GetFunction("gaus")->Draw();
```

Retrieve a pointer to the fitted TF1

The ponter can be used to manipulate the function: plotting, changing style, inspecting parameters...

```
TFitResultPtr fitres = myh2->Fit("gaus","S");

fitres->Parameter(2);
fitres->ParError(2);
fitres->Chi2();
fitres->Ndf();
```

- Some predefined fit functions available:
  - Exponential
  - Gaussian
  - Landau
  - Polynomial (up to 9th degree)
- Most common options:
  - "L" likelihood fit
  - "WL" likelihood fit with weighed points
  - "Q" quiet mode: do not print info on screen
  - "S" save the results in a TFitResultPtr opbect.

Stopnie swobody

Stopnie swobody

- Wariancję populacji określamy przez sumę kwadratów różnic:

$$S(X)=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}X_i\right)^2\right) \qquad \bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

Dlaczego $1/(n\text{-}1)$? To jest średnia różnicy kwadratów, powinno być $1/n$

- Zajmijmy się kwadratami różnic: $\sum_{i=1}^{n}(X_i-\bar{X})^2$

- Wartości X_i mogą przybierać <u>dowolne wartości</u>

- Dowolność ta jest jednak ograniczona ograniczony warunkiem (**wiązanie**) istnienia wartości średniej:

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \Rightarrow \boxed{\sum_{i=1}^{n}(X_i-\bar{X})=0}$$

- Mówimy, że **liczba stopni swobody dla sumy kwadratów wynosi** n-1

- Suma kwadratów dzielona przez liczbę stopni swobody to **odchylenie średnie kwadratowe (root-mean square - RMS)**

Pobieranie próby z rozkładów cząstkowych

Próby z rozkładów cząstkowych

- Czasem mamy do czynienia z sytuacją, gdy nie da się wybrać losowo próby z populacji generalnej
- Możemy jednak podzielić populację generalną na **podpopulacje**:
- Przykład:
  - badamy jakąś cechę wszystkich studentów w Europie – populacja generalna
  - najłatwiej to zrobić poprzez badanie tej cechy na poszczególnych uniwersytetach – podpopulacje
  - rozkład podpopulacji **nie jest** taki sam jak rozkład populacji generalnej
  - podpopulacje są jednak związane z całą populacją
- W jaki sposób wnioskować o całej populacji na podstawie prób losowych wybranych z podpopulacji?

Próby z rozkładów cząstkowych

- Dzielimy populację generalną G na t podpolujacji G_i
- Prawdopodobieństwo, że zmienna losowa należy do Gi: $P(X \in G_i) = p_i$
- Podpopulacje G_i, które są opisane gęstościami prawdopodobieństwa $f_i(x)$ i mają odpowiednie dystrybuanty $F_i(x)$:

$$F_i(x) = \int_{-\infty}^{x} f_i(x)\,dx = P(X \leqslant x | x \in G_i)$$

- Dla całej populacji mamy zatem:

$$F(x) = P(X \leqslant x | x \in G) = \sum_{i=1}^{t} P(X < x | X \in G_i) P(X \in G_i) = \sum_{i=1}^{t} P(X \in G_i) F_i(x) = \sum_{i=1}^{t} p_i F_i(x)$$

- Dla gęstości prawdopodobieństwa populacji: $f(x) = \sum_{i=1}^{t} p_i f_i(x)$

- Obliczamy wartość oczekiwaną populacji G:

$$\hat{x} = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \sum_{i=1}^{t} p_i \int_{-\infty}^{\infty} x f_i(x)\,dx = \boxed{\sum_{i=1}^{t} p_i \hat{x}_i}$$

Próby z rozkładów cząstkowych

- Wartość oczekiwana populacji G:

$$\hat{x}=E(X)=\int_{-\infty}^{\infty} xf(x)dx=\sum_{i=1}^{t} p_i \int_{-\infty}^{\infty} xf_i(x)dx=\boxed{\sum_{i=1}^{t} p_i \hat{x}_i}$$

- **Wniosek:** wartość oczekiwana z populacji to średnia ważona wartości oczekiwanych podpopulacji pomnożonych przez ich prawdopodobieństwa
- Jak zatem wyznaczyć **wariancję populacji** na podstawie podpopulacji?

$$\sigma^2(X)=E\left(\left(X-\hat{x}\right)^2\right)=\sum_{i=1}^{t} p_i E\left(\left[\left(X-\hat{x}_i\right)+\left(\hat{x}_i-\hat{x}\right)\right]^2\right)=\boxed{\sum_{i=1}^{t} p_i\left(\sigma_i^2+\left(\hat{x}_i-\hat{x}\right)^2\right)}$$

- **Wniosek:** wariancja populacji jest <u>średnią ważoną</u> wariancji z podpopulacji σ_i i wariancji wartości średniej podpopulacji $\hat{x}_i$ względem wartości średniej z całej populacji $\hat{x}$
- To nie koniec… Teraz musimy wybrać **próby losowe z podpopulacji** i **policzyć estymatory**

Próby z rozkładów cząstkowych

- Teraz z każdej podpopulacji wybierzmy próbkę o liczności n_i, w sumie n elementów: $n=\sum_{n=1}^{t} n_i$. Średnia arytmetyczna z całej próby wynosi wtedy:

$$\bar{X}_p = \frac{1}{n} \sum_{i=1}^{t} \sum_{j=1}^{n_i} X_{ij} = \frac{1}{n} \sum_{i=1}^{t} n_i \bar{X}_i$$

- Wartość oczekiwana i wariancja (niepewność) średniej z <u>całej próby</u>:

$$E(\bar{X}_p) = \frac{1}{n} \sum_{i=1}^{t} n_i \hat{x}_i$$

próbki z poszczególnych podpopulacji

$$\sigma^2(\bar{X}_p) = \frac{1}{n^2} \sum_{i=1}^{t} n_i^2 E\left((\bar{X}_i - \hat{x}_i)^2\right) = \frac{1}{n^2} \sum_{i=1}^{t} n_i^2 \sigma^2(\bar{X}_i) = \frac{1}{n} \sum_{i=1}^{t} \frac{n_i}{n} \sigma_i^2$$

Estymatory dla rozkładów cząstkowych

- Zauważmy jednak, że wartość średnia $\bar{X}_p$ <u>nie może</u> być estymatorem wartości oczekiwanej z całej populacji $\hat{x}$, gdyż zależy ona od dowolnego wyboru wielkości n_i próbek cząstkowych

$$\boxed{E\left(S\left(X_1, X_2, \ldots, X_n\right)\right) = \lambda, \text{ dla każdego } n}$$

- Jeśli jednak porównany wzory na średnią z populacji i całej próby:

$$\hat{x} = \sum_{i=1}^{t} p_i \hat{x}_i \qquad \bar{X}_p = \frac{1}{n} \sum_{i=1}^{t} n_i \hat{x}_i$$

- To widać, że jeżeli warunek $p_i = n_i / n$ jest spełniony, to wartość średnia z populacji $\hat{x}$ może być estymowana przez wyznaczenie najpierw wartości średnich poszczególnych prób $\bar{X}_i$, wewnątrz poszczególnych podpopulacji, a potem przez wyrażenie:

$$\tilde{X} = \sum_{i=1}^{t} p_i \bar{X}_i$$

- estymator wartości średniej z populacji zależeny od wartości średnich z prób

$$p_i = \frac{n_i}{n}$$

- Wariancja powyższego estymatora: $$\sigma^2\left(\tilde{X}\right) = \sum_{i=1}^{t} p_i^2 \sigma^2\left(\bar{X}_i\right) = \sum_{i=1}^{t} \frac{p_i^2}{n_i} \sigma_i^2$$

- Oczywiście, chcielibyśmy tak dobierać próbki, by była ona jak najmniejsza: $$n_i = n\, p_i \sigma_i / \sum_{i=1}^{t} p_i \sigma_i$$

Estymatory dla rozkładów cząstkowych

- Gdzie my to wszystko wykorzystujemy i po co?
  - Możemy sobie wyobrazić badania społeczne, gdzie próbujemy wnioskować na temat całej populacji poprzez analizy poszczególnych podgrup (podpopulacji) – np. analizujemy dane na temat wszystkich studentów w Europie poprzez poszczególne analizy studentów poszczególnych typów uczelni (np. osobno techniczne, medyczne, ogólne)
  - Albo rozkład powierzchni zajmowanej przez pewien gatunek trawy
  - Czy przeprowadzamy badania kliniczne leku w wielu ośrodkach w różnych krajach

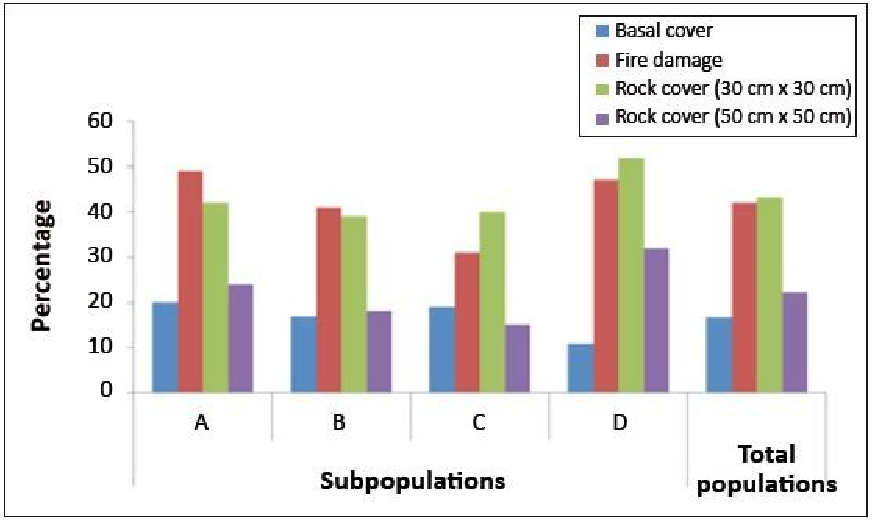


FIGURE 6: Percentage basal grass cover, fire damage and rock cover per subpopulation and total population of *Haworthia koelmaniorum* var. *Mcmurtryi* within the study area.

Rozkład χ2

Rozkład χ2

- Załóżmy (dla uproszczenia), że badamy populację opisaną standardowym rozkładem Gaussa o wartości oczekiwanej 0 i wariancji 1
- Pobieramy z niego próbę n-elementową i tworzymy sumę kwadratów:

$$\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

- Zmienna χ^2 podlega **rozkładowi chi-kwadrat** o n stopniach swobody
- Funkcja gęstości prawdopodobieństwa:

$$f(\chi^2) = k \cdot (\chi^2)^{\lambda - 1} e^{-1/2 \chi^2} \qquad k = \frac{1}{\Gamma(\lambda) 2^\lambda}$$

- Dystrybuanta:

$$F(\chi^2) = \frac{1}{\Gamma(\lambda) 2^\lambda} \int_0^{\chi^2} u^{\lambda - 1} e^{-1/2 u} du$$

  – gdzie $\lambda = 1/2 n$ a n to liczba stopni swobody

Rozkład χ2

- Funkcja charakterystyczna rozkładu χ^2 ma postać:

  $$\phi_{\chi^2}(t) = (1-2it)^{-\lambda}$$

- Korzystając z jej własności funkcji charakterystycznej otrzymujemy natychmiast, że suma dwóch różnych χ^2 o n_1 i n_2 stopniach swobody daje również rozkład χ^2 o $n=n_1+n_2$ stopniach swobody
- Różniczkując funkcję charakterystyczną możemy wyznaczyć wartość oczekiwaną i wariancję rozkładu χ^2:

  $$E\{X^2\} = -i\phi_{\chi^2}{}'(0) = 2\lambda \equiv n$$

  $$E\{(X^2)^2\} = -i\phi_{\chi^2}{}''(0) = 4\lambda^2 + 4\lambda$$

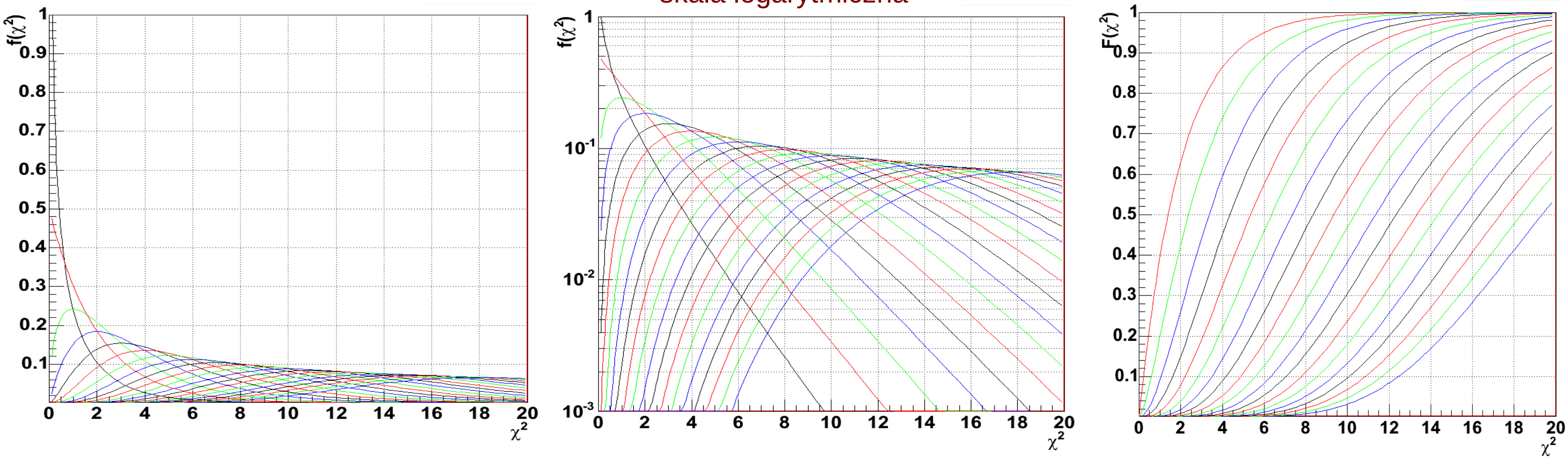
  $$\sigma^2(X^2) = E\{(X^2)^2\} - (E\{X^2\})^2 = 4\lambda \equiv 2n$$

- Czyli wartość wartość oczekinawa rozkładu χ^2 wynosi n a wariancja $2n$

Rozkład χ2

- Wykresy rozkładu i dystrybuanty rozkładu χ^2 dla n od 1 do 20

skala logarytmiczna



- W rzeczywistych przypadkach mamy do czynienia z pełnym rozkładem Gaussa o dowolnym a i σ. Wprowadzamy wtedy odpowiednie przeskalowanie

$$\chi^2 = X^2 = \frac{(X_1 - a)^2 + (X_2 - a)^2 + \ldots + (X_n - a)^2}{\sigma^2}$$

- a w ogólnym przypadku wielowymiarowym, gdy zmienne są zależne:

$$\chi^2 = X^2 = (\boldsymbol{X} - \boldsymbol{a})^T B (\boldsymbol{X} - \boldsymbol{a})$$

Rozkład t-Studenta

Rozkład t-Studenta

- Załóżmy $X_1,\ldots, X_n$ są zmiennymi losowymi pochodzącymi z rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$
- Estymator wartości oczekiwanej (średnia z próby) to:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  a estymator wariancji to:

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

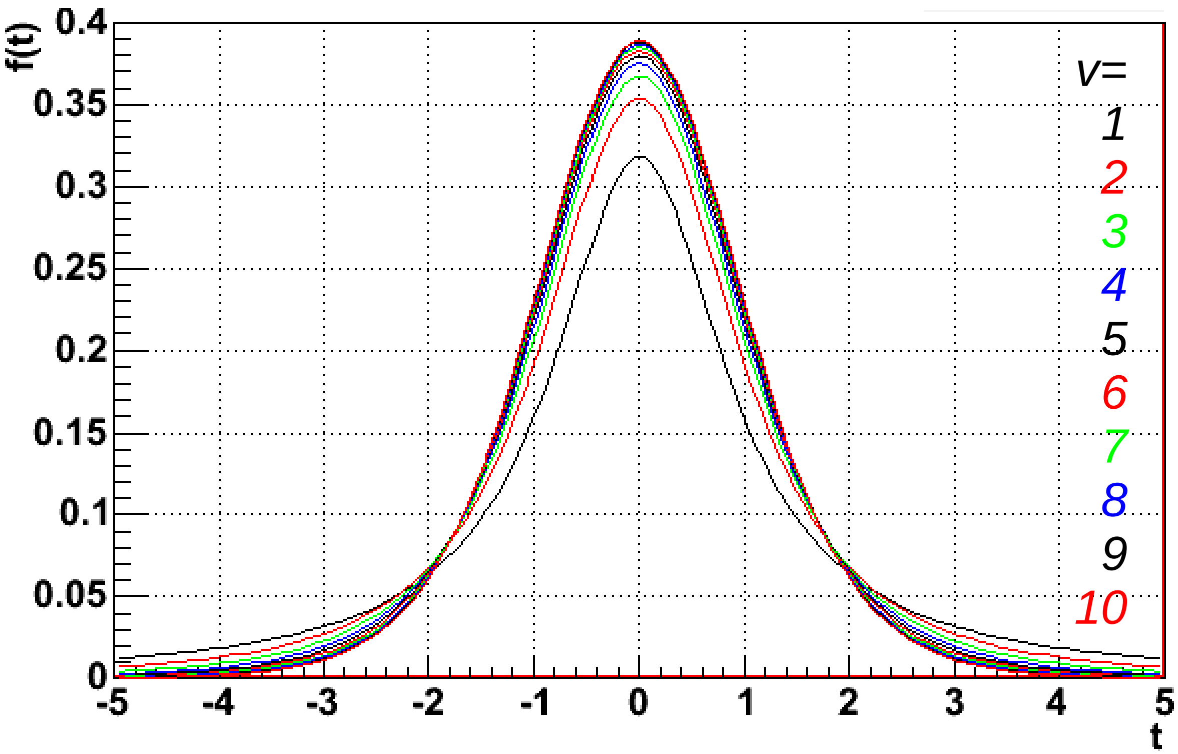
- Wtedy, zmienna $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ ma rozkład normalny standardowy (o średniej 1 i odchyleniu 0)
- A zmienna $\frac{\bar{X}-\mu}{S/\sqrt{n}}$, gdzie zastąpiliśmy odchylenie jego estymatorem ma **rozkład t-Studenta** o n-1 stopniach swobody

Rozkład t-Studenta

- Funkcja gęstości prawdopodobieństwa

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

ν– liczba stopni swobody

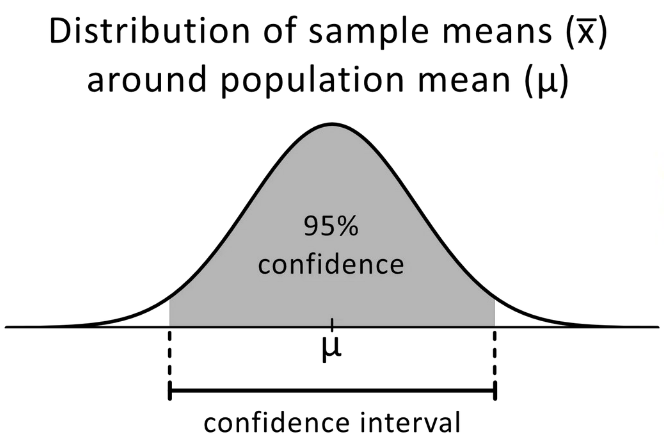


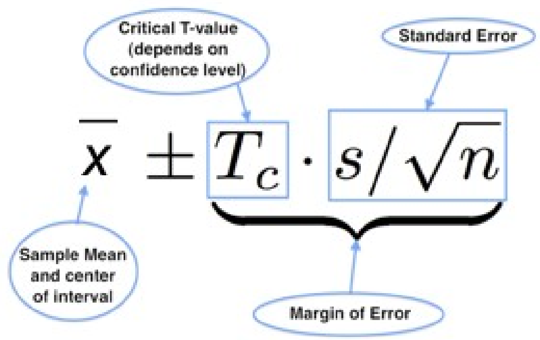
- Jak widać, rozkład t-Studenta jest symetryczny względem 0
- Rozkład **dąży do rozkładu Gaussa gdy** $\nu \to \infty$
- Z symetrii względem 0 mamy związek (analogicznie jak dla Gaussa): $P(|t| \leq t) = 2F(|t|) - 1$

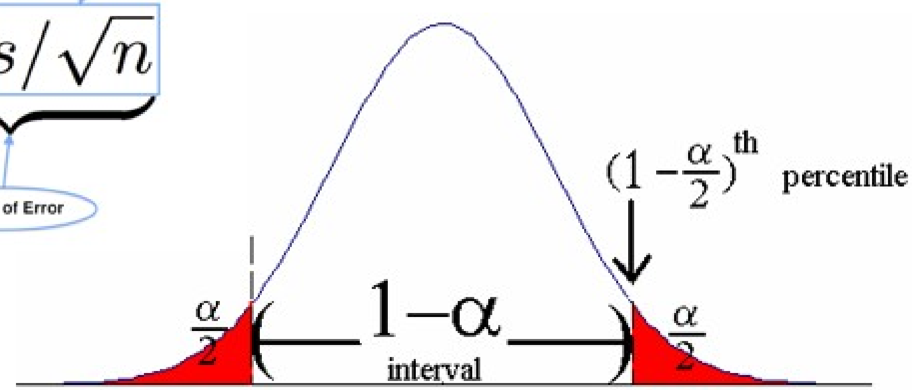
Poziom ufności

Poziom ufności

- Rozkłady χ^2 oraz t-Studenta pozwalają nam wprowadzić pojęcie **poziomu ufności** dla estymowanych wartości średnich oraz wariancji
- **Wartość średnia z próby**
  - pokazaliśmy, że rozkład t-Studenta ma wielkość $\frac{\bar{X}-\mu}{S/\sqrt{n}}$
  - możemy zatem zdefiniować prawdopodobieństwo wokoł wartości oczekiwanej

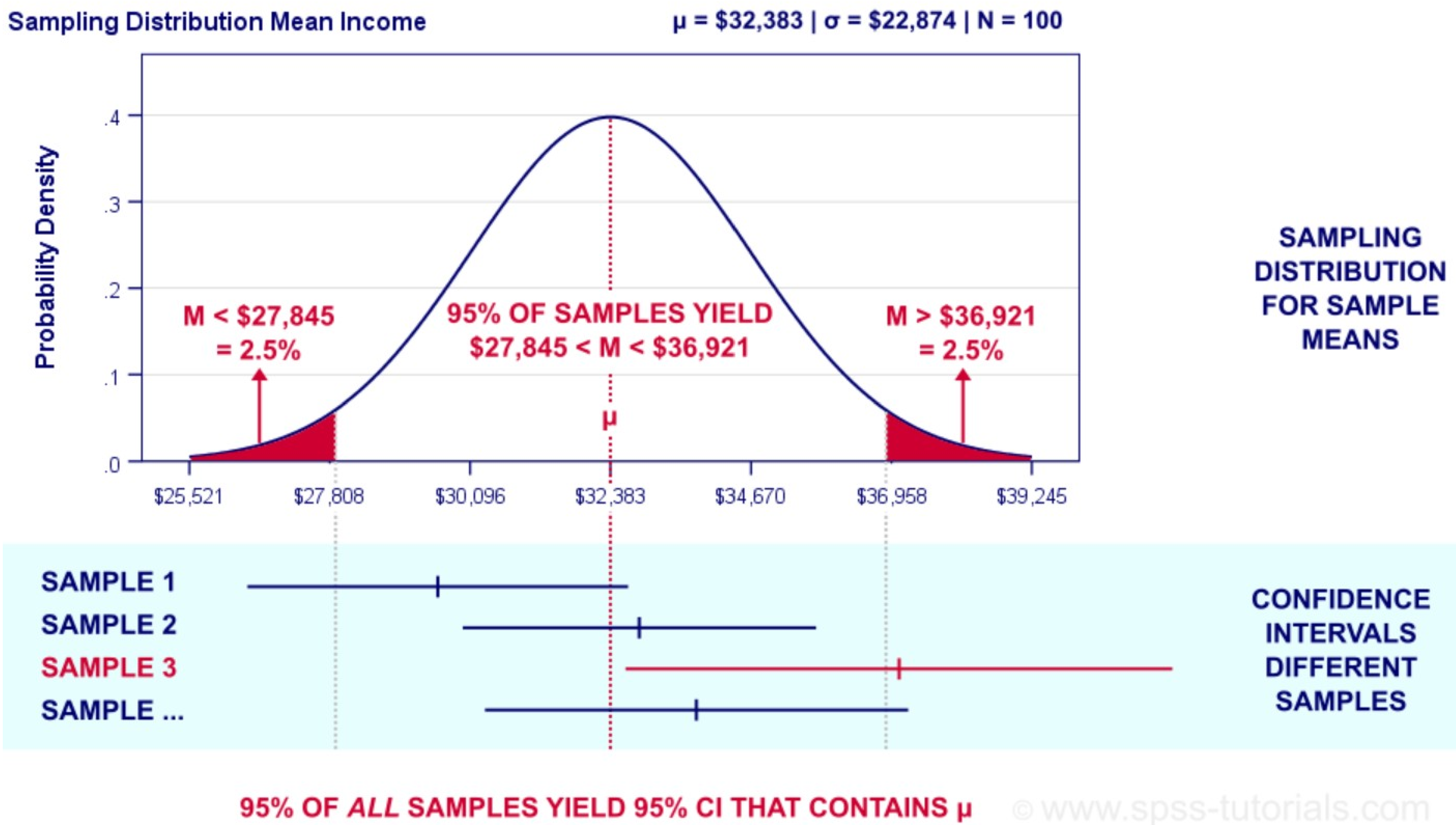






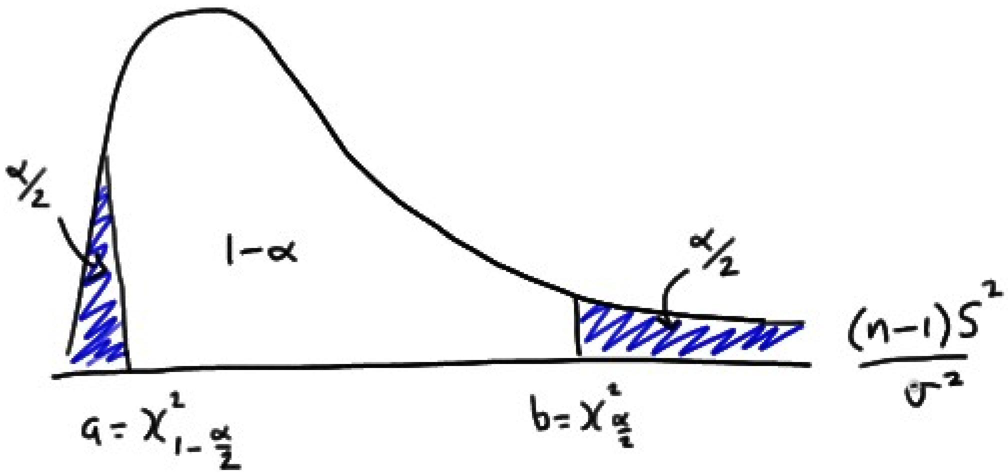
  - oczekujemy, że nasz wynik (estymowana wartość średnia z próby) mieści się w przedziale wyznaczonym przez poziom ufności
  - Dla dużych wartości n możemy to prawdopodobieństwo estymować rozkładem normalnym (t-Student → rozkład normalny)

Poziom ufności



Poziom ufności

- Rozkłady χ^2 oraz t-Studenta pozwalają nam wprowadzić pojęcie **poziomu ufności** dla estymowanych wartości średnich oraz wariancji
- **Wariancja z próby**
  - Można, że rozkład χ^2 ma wielkość $\frac{n-1}{\sigma^2} S^2$ estymatyor wariancji
  - możemy zatem zdefiniować prawdopodobieństwo wokół wariancji



$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

  - oczekujemy, że nasz wynik (estymowana wariancja z próby) mieści się w przedziale wyznaczonym przez poziom ufności

https://en.wikipedia.org/wiki/Confidence_interval

Weryfikacja hipotez statystycznych

Weryfikacja hipotez statystycznych

- **Przykład:** rozważamy zmienną losową X opisaną standardowym rozkładem Gaussa (średnia 0, odchylenie 1). Pobieramy 10-elementową próbę, uzyskaliśmy średnią arytmetyczną: $\bar{X}=0{,}5$
- Jak na podstawie tej jednej realizacji próby (np. wyniku eksperymentu) możemy stwierdzić, czy pochodzi ona z takiej populacji? Innymi słowy, naszą **hipotezą** jest: **próba losowa pochodzi z rozkładu Gaussa o średniej 0 i odchyleniu 1**
- Procedura weryfikacji hipotezy nazywana jest **testem statystycznym**
- **Jeżeli hipoteza jest słuszna (nasze założenie)** to wartość średnia (będąca również zmienną losową) $\bar{X}$ ma rozkład normalny ze średnią 0 i odchyleniem std. $1/\sqrt{10}$

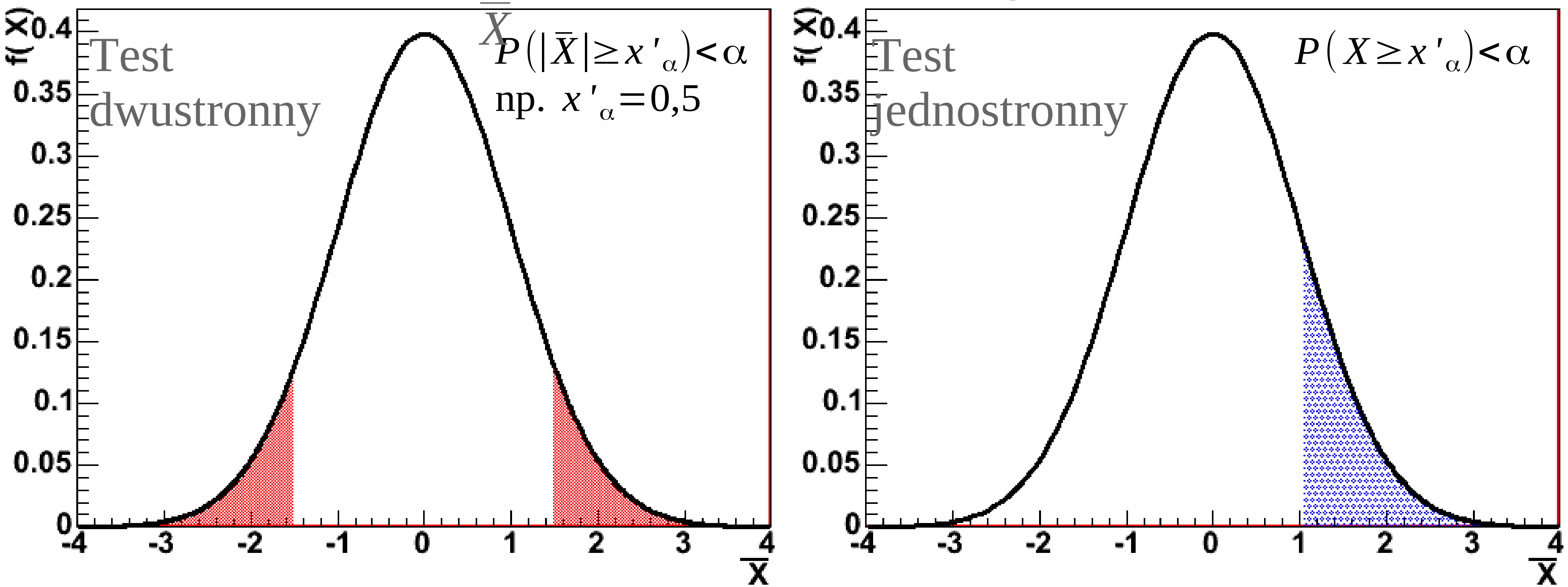
$$\sigma^2(\bar{X})=\frac{1}{n}\sigma^2(X)=\frac{1}{10}\cdot 1\Rightarrow\sqrt{\sigma^2(\bar{X})}=\frac{1}{\sqrt{10}}$$

Weryfikacja hipotez statystycznych

- Jak na podstawie **konkretnej realizacji próby** sprawdzić, czy założona hipoteza jest prawdziwa?
  - **I:** musimy ustalić pewną wartość prawdopodobieństwa α (zwanego **poziomem istotności**,z reguły mała wartość, np. 0,01, albo 0,03, czy 0,05)
  - **II:** pytamy, czy prawdopodobieństwo zaobserwowania określonych wartości próby jest mniejsze niż α: $P(|\bar{X}| \geq 0{,}5) < \alpha$
  - **nierówność spełniona** – jest mało prawdopodobne, aby próba pochodziła z rozkładu określonego przez testowaną hipotezę → **możemy ją odrzucić**
  - **prawdopodobieństwo zaobserwowania tego, że $|\bar{X}|$ jest duże, jest bardzo małe, ale takie nam się trafiło – więc prawdopodobnie (z prawdodobieństwem $1\text{-}\alpha$) nasza hipoteza nie jest słuszna**
  - **III:** jeśli prawdopodobieństwo jest mniejsze niż przyjęta wartość prawdopodobieństwa (poziom istotności) α, odrzucamy hipotezę na zadanym poziomie istotności

Weryfikacja hipotez statystycznych

Rozkład wartości średniej $\bar{X}$



- Jeśli (w naszym przykładzie) wartość średnia znajduje się w zaznaczonym obszarze (nazywamy go **obszarem krytycznym**), to hipotezę odrzucamy
  - jeśli oczekujemy rozkładu normalnego o średniej 0 i małym odchyleniu (np. 10), a z próby losowej (konkretny eksperyment) mamy średnią 1000, to lądujemy w **"ogonie"** rozkładu średniej i na podstawie tej konkretnej próby **odrzucamy** hipotezę (**ale na podstawie innej próby moglibyśmy zaakceptować**)

Weryfikacja hipotez statystycznych

- W ogólnym przypadku używamy innych wielkości niż średnia:
  - definiujemy jakąś (wygodną dla nas) statystykę testową T (np. różnicę między wynikiem eksperymentu a krzywą teoretyczną)
  - ustalamy poziom istotności α
  - wyznaczamy taki zbiór U, który określa obszar zmienności statystyki testowej T, taki że prawdopodobieństwo znalezienia się w nim jest ograniczone wartością α: $P(T \in U) = \alpha$
  - z pobranej próby wyznaczamy konkretną wartość statystyki testowej T': jeżeli znajduje się ona **wewnątrz** obszaru krytycznego U, **odrzucamy hipotezę** (mówimy: krzywa teoretyczna nie opisuje wyniku eksperymentu), czyli odrzucamy hipotezę, jeżeli $T' \in U$

Hipoteza zerowa a hipoteza alternatywna

- Każde testowanie hipotez zaczynamy od założenia, że **nie ma związku** między zmiennymi lub nie ma różnic między porównywanymi grupami.

Hipoteza zerowa (H_0) zawsze mówi o braku związku między zmiennymi lub o braku różnicy.

- Hipoteza alternatywna (badawcza), H_A, stawiana jest przez nas i dotyczy sytuacji wytłumaczenia zjawiska, kiedy hipoteza zerowa jest fałszywa
  - z metodologicznego punktu widzenia nie jest możliwe, aby w pełni udowodnić jej prawdziwość.
  - wystarczy, ze pojawi się jeden przypadek zaprzeczający hipotezie i staje się ona fałszywa (odnosi się to również do teorii).

Hipotezy w badaniach falsyfikujemy a nie potwierdzamy.

Hipoteza zerowa i alternatywna

- Hipoteza może być różna, przykłady hipotez:
  - średni wzrost Polaków to 175 cm
  - 2% dzieci w wieku szkolnym nie lubi czekolady
  - poziom szczęścia dwie minuty po zjedzeniu dużej porcji lodów jest wyższy niż przed zjedzeniem tejże dużej porcji lodów
- **Hipoteza zerowa** (H_0) to w uproszczeniu taka, gdy nie widzimy różnicy – np. czujemy się tak samo po zjedzeniu lodów jak przed
  - w przypadku pomiarów, np. wartość $\chi 2$ jest mała → teoria opisuje dane
- **Hipoteza alternatywa** (H_A) to przeciwieństwo hipotezy zerowej, którą <u>możemy zdefiniować na kilka sposobów</u>, np.:
  - np. jest różnica w szczęściu w zjedzeniu lodów (test dwustronny)
  - poziom szczęścia po zjedzeniu lodów jest mniejszy (test jednostronny)
  - poziom szczęścia po zjedzeniu lodów jest większy (test jednostronny)
  - w przypadku pomiarów $\chi 2$ jest duże → teoria nie opisuje danych (test jednostronny – rozkład $\chi 2$ jest niesymetryczny)
- **Statystyka testowa** - to funkcja próby, na podstawie której wnioskuje się o odrzuceniu lub nie hipotezy statystycznej – wielkość mająca swój rozkład prawd.

Test dobroci χ^2 dopasowania

Test χ^2 dobroci dopasowania

- Mamy N pomiarów g_i, $i=1, 2, \ldots, N$ oraz ich niepewności σ_i
- Wartości $f_i,\ i=1,2,\ldots,N$ określają nam prawdziwy rozkład danej wielkości mierzonej (**np. znaleziony poprzez estymację**)
- **Dla każdego pomiaru liczymy wielkość** u_i**:** $u_i=\dfrac{g_i-f_i}{\sigma_i},\ i=1,2,\ldots,N$
- Jeśli nasza teoria (wartości f_i) jest prawdziwa, to <u>rozkłady różnic</u> u_i <u>mają postać standardowego rozkładu normalnego</u> **– nasza hipoteza**
- Jeśli tak, to rozkład χ^2 o N stopniach swobody będzie miała wielkość:
$$T=\sum_{i=1}^{N} u_i^2=\sum_{i=1}^{N}\left(\frac{g_i-f_i}{\sigma_i}\right)^2$$
- **(Subiektywnie)** oczekujemy małej wartości wielkości T
- Gdy hipoteza jest **fałszywa**, wówczas poszczególne różnice u_i przyjmują duże wartości (wartość T jest duża)
- Jak określić granicę zmienności T? Można zauważyć, że granica ta jest określona **kwantylem** $\chi^2_{1-\alpha}$, czyli: $\boxed{P(T>\chi^2_{1-\alpha})=\alpha}$

$$\boxed{\begin{aligned}F(x_q)=P(X\le x_q)&=q\\ P(X>x_q)&=1-q\end{aligned}}$$

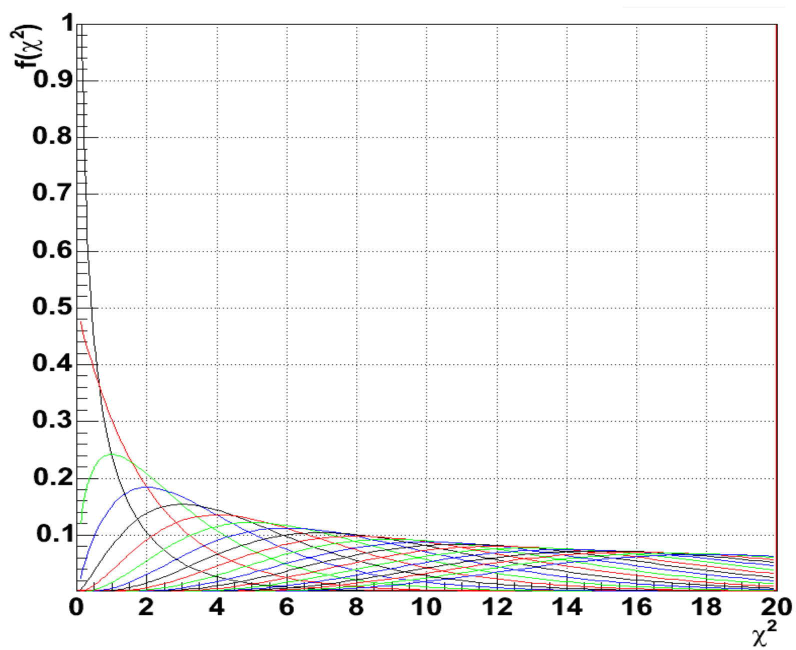
Test χ^2 dobroci dopasowania

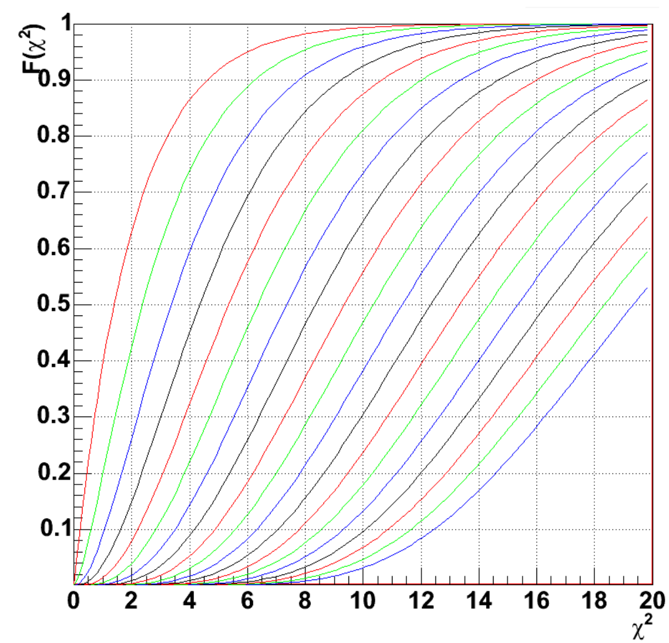
- **<u>Podsumowując</u>, w naszym przypadku musimy <u>dla danej realizacji próby (wyniku eksperymentu)</u> wyznaczyć wartość testową T i porównać ją z odpowiednim kwantylem rozkładu** χ^2 o **odpowiedniej liczbie stopni swobody:**

$$\boxed{T > \chi^2_{1-\alpha}}$$

- **Jeżeli ten warunek jest spełniony, to hipotezę odrzucamy** (punkty teoretyczne nie opisują danych eksperymentalnych na zadanym poziomie istotności)

- Skąd wziąć kwantyl? Z tablic lub z dystrybuanty:





Test χ^2 i doświadczalny rozkład częstości

- Możemy również rozważać zmienną losową X, (opisaną rozkładem $f(x)$) którą dzielimy na r przedziałów (histogram): $\xi_1, \xi_2, \ldots, \xi_i, \ldots, \xi_r$
- Całkując $f(x)$ w przedziałach otrzymujemy prawdopodobieństwo p_i zaobserwowania zmiennej X w danym przedziale (binie):

$$p_i = P(x \in \xi_i) = \int_{\xi_i} f(x)\,dx\,; \quad \sum_{i=1}^{r} p_i = 1$$

- Z pobranej próby o liczebności n oznaczamy przez n_i elementy leżące w danym przedziale ξ_i
- Oczywiście zachodzi relacja: $n = \sum_{i=1}^{r} n_i$
  suma wejść w poszczególnych binach równa jest liczebności próby
- **Oczekiwalibyśmy** (zakładając prawdziwość $f(x)$)**, że:** $n_i = np_i$
- **Hipoteza:** zakładamy, że dla dużych wartości liczb n_i ich wariancja równa się n_i (patrz dyskusja o rozkładzie Poissona) i że rozkład wielkości u_i: $u_i^2 = \frac{(n_i - n\,p_i)^2}{n_i}, \ \ lub \ \ u_i^2 = \frac{(n_i - n\,p_i)^2}{np_i}$ ma rozkład Gaussa

Test χ^2 i doświadczalny rozkład częstości

- Wtedy, suma kwadratów: $T=\sum_{i=1}^{r} u_i^2$
- Będzie miała (dla dużych n) rozkład χ^2
- **Jaka jest liczba stopni swobody?** Z defninicji histogramu mamy jedno **równanie więzów**: $n=\sum_{i=1}^{r} n_i$
- Zatem zmienne u_i **nie są niezależne**, więc liczba stopni swobody równa się r-1
- Oczywiście, jeżeli dodatkowo estymujemy p parametrów rozkładu na podstawie pomiarów (wprowadzamy p kolejnych wiezów uzależniających od siebie wielkości u_i), to liczba stopni swobody wynosi r-1-p
- Wartość T porównujemy, tak jak do tej pory, z kwantylami rozkładu χ^2 o określonej liczbie stopni swobody dla zadanego poziomu istotności α: $T>\chi^2_{1-\alpha}$
- **Jeśli nierówność jest spełniona – odrzucamy hipotezę**

Test χ^2 - przykład

Zadanie

Weryfikacja hipotez statystycznych (5 pkt.)

- Przeprowadzono eksperyment naświetlania wodorowej komory pęcherzykowej wiązką fotonów w celu badania oddziaływań fotonów z protonami. Fotony powodują powstawanie par elektron-pozyton, które mogą być wykorzystane do monitorowania wiązki fotonów. Częstość występowania zdjęć z 0,1,2,... parami elektron-pozyton powinna podlegać rozkładowi Poissona. Należy wczytać dane z pliku (w pierwszej kolumnie znajduje się liczba par elektronowych na zdjęciu k, a w drugiej liczba zdjęć zawierających k par elektronowych). Widzimy, że rozkład ten przypomina rozkład Poissona - próbujemy zatem obliczyć estymator największej wiarygodności dla parametry rozkładu Poissona (patrz Wykład 10 slajd 13) (1 pkt.)
- Narysować na jednym wykresie punkty pomiarowe i dopasowanie (metodą estymatora największej wiarygodności).
- Sprawdzić jakość dopasowania za pomocą testu χ2. W tym celu należy zaimplementować funkcję obliczającą statystykę testową χ2 zgodnie z wzorem $T = \sum_k \frac{(n_k - np_k)^2}{np_k}$

gdzie: nk - liczba obserwacji w k-tym binie, npk - przewidywana przez teorię liczba przypadków w k-tym binie

- Okreslić liczbę stopni swobody i obliczyć wartość statystyki testowej. (1 pkt.)
- Zaimplementować funkcję zwracającą wynik testu χ2 na zadanym poziomie istotności α

Wykorzystując zaimplementowaną funkcję zweryfikować hipotezę mówiacą, że dane pomiarowe podlegają rozkładowi Poissona. Dobrać odpowiednią wartość poziomu istotności. Uwaga! Kwanyl możemy odczytać z policzonej na ostatnich zajęciach dystrybuanty. (2 pkt.)

Test χ^2 - przykład

Zadanie

Weryfikacja hipotez statystycznych (5 pkt.)

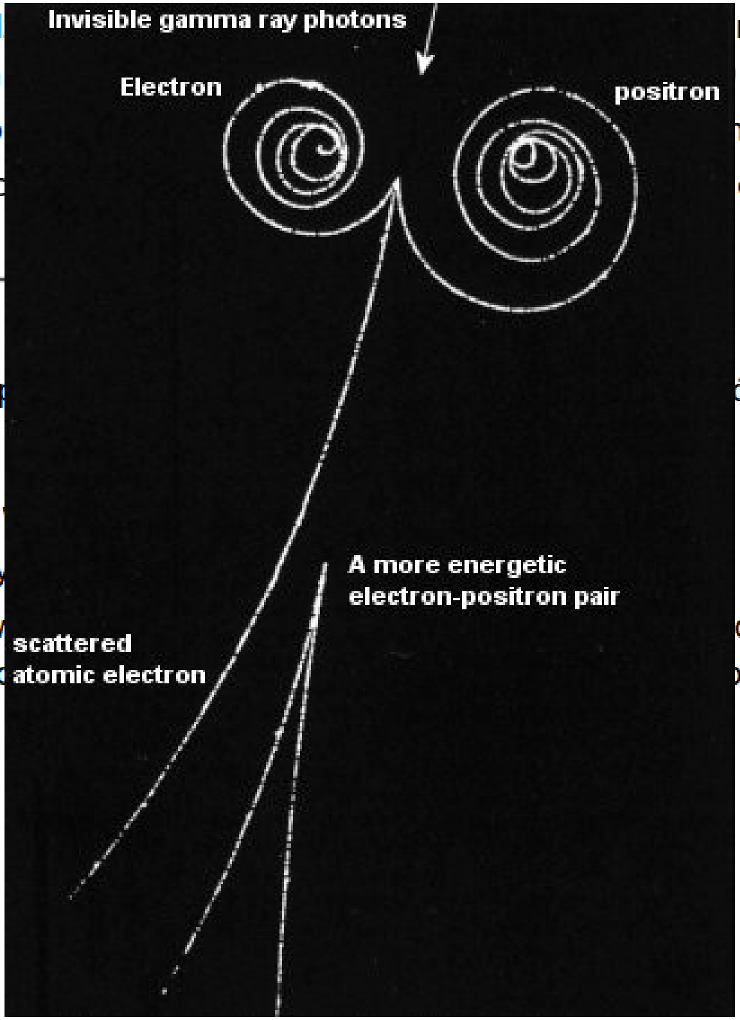
- Przeprowadzono eksperyment naświetlania wodorowej komory pęcherzykowej wiązką fotonów w celu badania oddziaływań fotonów z protonami. Fotony powodują powstawanie par elektron-pozyton, które mogą być wykorzystane do monitorowania wiązki fotonów. Częstość występowania zdjęć z 0,1,2,... parami elektron-pozyton powinna podlegać rozkładowi Poissona. Należy wczytać dane z pliku (w pierwszej kolumnie znajduje się liczba par elektronowych na zdjęciu k, a w drugiej liczba zdjęć zawierających k par elektronowych). Wid … na - próbujemy zatem obliczyć estymator największej wiarygodności dla … slajd 13) (1 pkt.)
- Narysować na jednym wykresie punkty p … największej wiarygodności).
- Sprawdzić jakość dopasowania za pomoc … ć funkcję obliczającą statystykę testową

χ2 zgodnie z wzorem $T = \sum_k \frac{(n_k \ldots}{\ldots}$

gdzie: nk - liczba obserwacji w k-tym binie, np … ów w k-tym binie

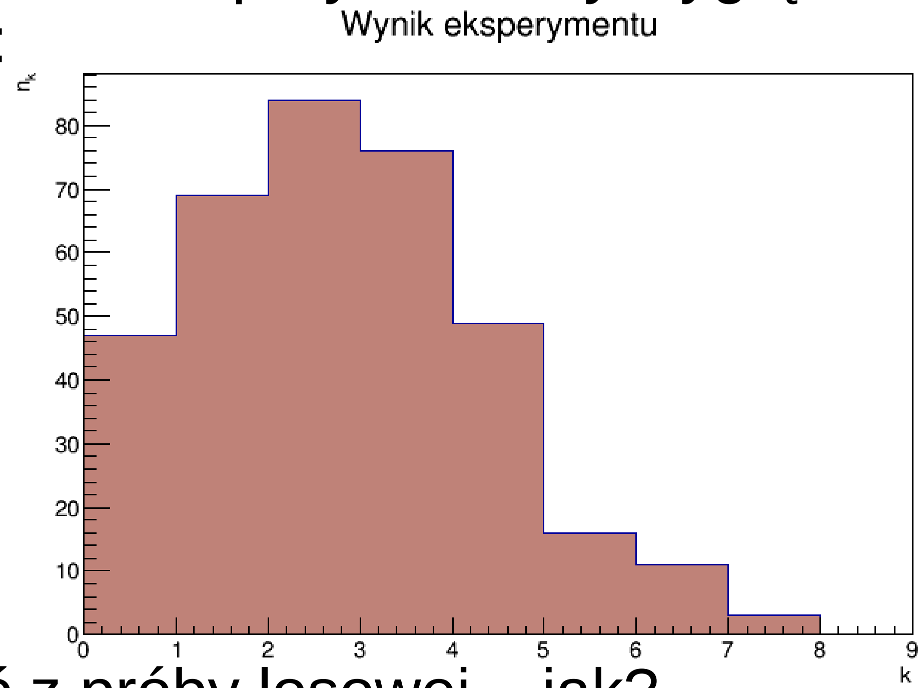
- Określić liczbę stopni swobody i obliczyć …
- Zaimplementować funkcję zwracającą wy …

Wykorzystując zaimplementowaną funkcję zw … owe podlegają rozkładowi Poissona. Dobrać odpowiednią wartość poziomu istotno … nej na ostatnich zajęciach dystrybuanty. (2 pkt.)



Test χ^2 - przykład

- Po wczytaniu danych z pliku histogram eksperymentalny wygląda następująco (nasza próba losowa):



- **Zakładamy hipotezę:** teoria mówi to jest rozkład Poissona („na oko” zresztą tak wygląda)
- Rozkład Poissona ma tylko jeden parametr (wartość średnią):

$$f(k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

- Musimy go zatem jakoś wyznaczyć z próby losowej – jak? **Na przykład metodą największej wiarygodności** – szukamy estymatora nieobciążonego największej wiarygodności o minimalnej wariancji – wyprowadziliśmy go sobie na Wykładzie 10:

$$\frac{dl}{d\lambda}=l'=\sum_{j=1}^{N}\left\{\frac{k^{(j)}}{\lambda}-1\right\}=$$

$$\frac{1}{\lambda}\sum_{j=1}^{N}\left\{k^{(j)}-\lambda\right\}=\boxed{\frac{N}{\lambda}(\bar{K}-\lambda)}$$

$$\boxed{\tilde{\lambda}=\bar{K},\ \ \sigma^2(\bar{K})=\frac{N}{\lambda}}$$

Przypomnienie – definicja estymatora o min. wariancji:

$$\boxed{\begin{array}{c} l'=A(\lambda)(\tilde{\lambda}-\lambda) \\ \sigma^2(S)=\dfrac{1}{|A(\lambda)|} \end{array}}$$

Test χ^2 - przykład

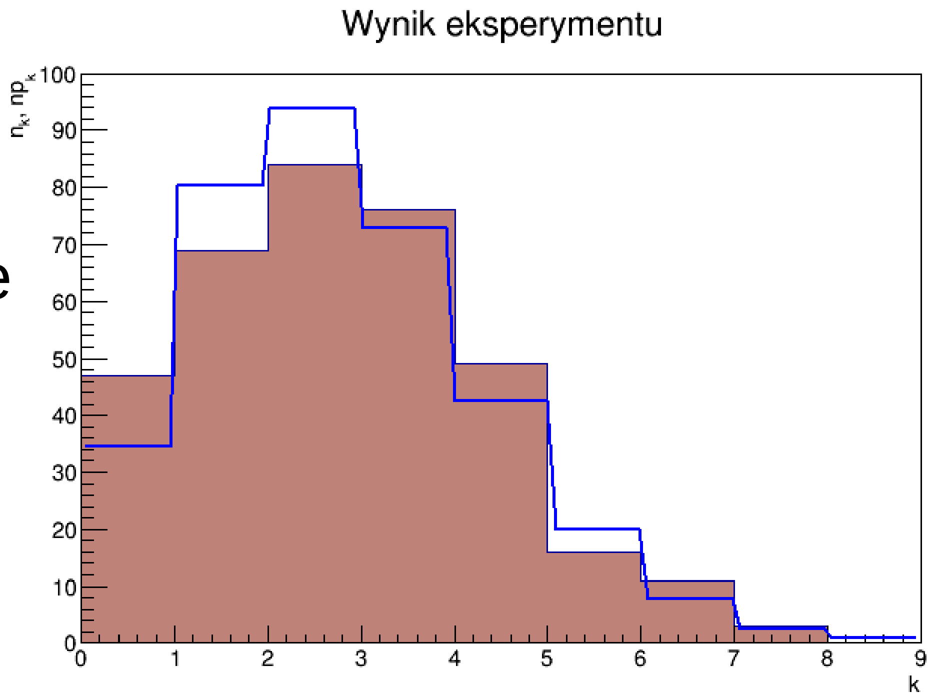
- Czyli estymatorem największej wiarygodności o minimalnej wariancji dla rozkładu Poissona jest średnia arytmetyczna z próby
- Oczywiście w naszym przypadku mamy histogram, który zawiera jakąś całkowitą liczbę wejść (całka z histogramu nie jest równa 1), wobec tego do średniej dodajemy wagi w postaci liczby wejść w danym binie i średnia staje się średnią ważoną:

$$\tilde{\lambda} = \frac{\sum_k k \cdot n_k}{\sum_k n_k}$$

- W naszym przypadku wartość ta wynosi mniej więcej: $\tilde{\lambda} \approx 2{,}33$
- Rysujemy więc funkcję:

$$n \cdot p_k = n \cdot f(k) = n \cdot \frac{\tilde{\lambda}^k}{k!} e^{-\tilde{\lambda}}, \text{gdzie } n = \sum_k^7 n_k$$

- Jak teraz sprawdzić, czy faktycznie nasza hipoteza jest słuszna?
- **Testujemy dobroć dopasowania**



Test χ^2 - przykład

- Musimy zatem wyznaczyć wartość statystyki testowej T:

$$T=\sum_{k=0}^{7} u_i^2=\sum_{k=0}^{7} \frac{(n_k-n\,p_k)^2}{np_k}\approx 10{,}53$$

- Co dalej? Zakładamy poziom istotności, na przykład: $\alpha=0{,}01$
- Musimy jeszcze określić liczbę stopni swobody – ile ich jest?
  - liczba binów (8) minus 1 minus liczba parametrów (1)

$$r-1-p=8-1-1=6$$

- Teraz szukamy odpowiedniego kwantyla rozkładu χ^2 o 6 stopniach swobody: $\chi^2_{1-\alpha}=\chi^2_{0{,}99}\approx 16{,}81$
- Porównujemy statystykę z kwantylem: $T=10{,}51<\chi^2_{0{,}99}=16{,}81$
- **Warunek** $T>\chi^2_{1-\alpha}$ **nie jest spełniony, zatem na poziomie istotności** $\alpha=0{,}01$ **nie ma podstaw do odrzucenia hipotezy**

Test równości wariancji (test F-Fishera)

Test równości wariancji (F-Fishera)

- **Problem:** porównywanie wariancji populacji o jednakowych wartościach średnich
- **Przykład:** pomiar tej samej wielkości dwoma przyrządami pomiarowymi (zakładamy brak niepewności systematycznych – typu B)
- **Pytanie (hipoteza):** czy pomiary będą miały jednakowe wariancje (czy dokładność pomiaru jest jednakowa dla obu przyrządów)?
- Załóżmy, że rozważane populacje mają rozkład normalny
- Pobieramy próby o liczebności N_1 i N_2
- Dla każdej z pobranych prób wyznaczamy wariancję i liczymy iloraz

$$F = s_1^2 / s_2^2 \qquad s^2(X) = \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - \bar{X} \right)^2 \qquad s^2(\bar{X}) = \frac{1}{N \cdot (N-1)} \sum_{i=1}^{N} \left(X_i - \bar{X} \right)^2 = s^2(X) / N$$

- Jeśli hipoteza o równości wariancji jest **prawdziwa**, to iloraz F powinien być bliski jedności $F \sim 1$

Test równości wariancji (F-Fishera)

- Można udowodnić, że tego typu wielkość ma rozkład F-Fishera

$$f(F)=\left(\frac{f_1}{f_2}\right)^{\frac{1}{2}f_1}\frac{\Gamma\left(\frac{1}{2}(f_1+f_2)\right)}{\Gamma\left(\frac{1}{2}f_1\right)\Gamma\left(\frac{1}{2}f_2\right)}F^{\frac{1}{2}f_1-1}\left(1+\frac{f_1}{f_2}F\right)^{-\frac{1}{2}(f_1+f_2)} \qquad f_1=N_1-1;\ \ f_2=N_2-1$$

- Szukamy zatem analogicznie wartości granicznej określającej obszar krytyczny, która jest odpowiednim kwantylem rozkładu F-Fischera:

$$P\left(\frac{s_1^2}{s_2^2}>F_{1-\alpha}\right)=\alpha$$

- Ostatecznie sprawdzamy zatem warunek: $\boxed{\frac{s_1^2}{s_2^2}>F_{1-\alpha}}$
- **To jest test jednostronny**, na ogół posługujemy jednak **testem dwustronnym**:

$$\boxed{\frac{s_g^2}{s_k^2}>F_{1-\alpha/2}(f_g,f_k),\ \text{gdzie f to liczby stopni swobody}}$$

Test równości wariancji (F-Fishera)

- Czyli w praktyce musimy zweryfikować hipotezę:

$$\frac{s_g^2}{s_k^2} > F_{1-\alpha/2}(f_g, f_k) = F''_{\alpha}(f_g, f_k)$$

- Indeksy g i k oznaczają większą i mniejszą wariancję z próby, czyli:

$$s_g^2 > s_k^2$$

- Jeżeli nierówność jest **spełniona**, to hipotezę o równości wariancji można **odrzucić**

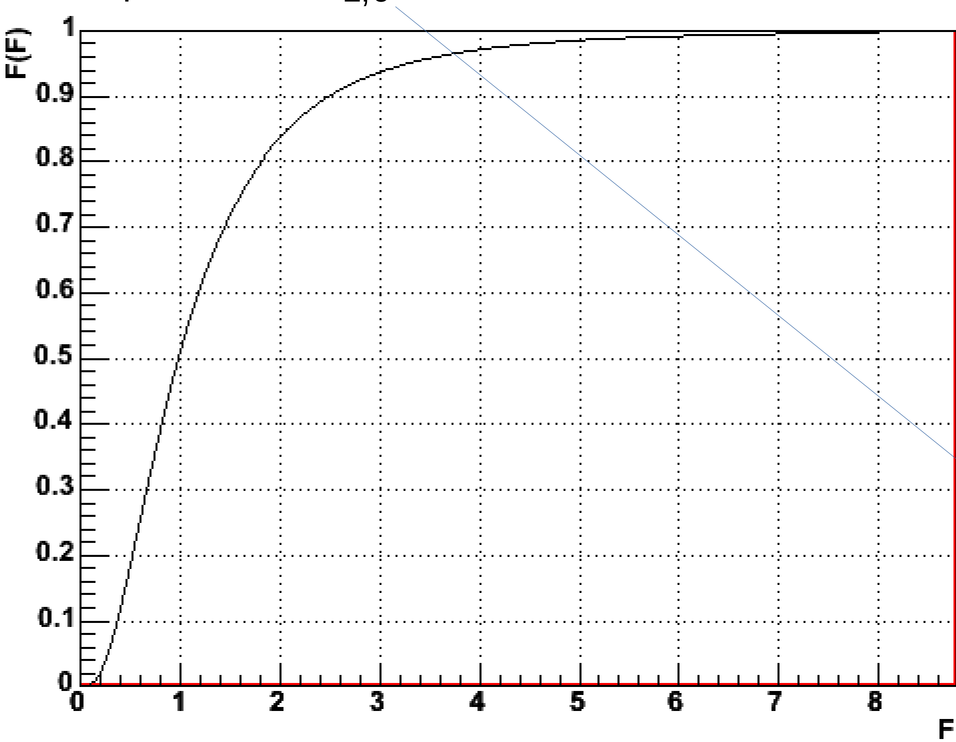
Test równości wariancji – przykład

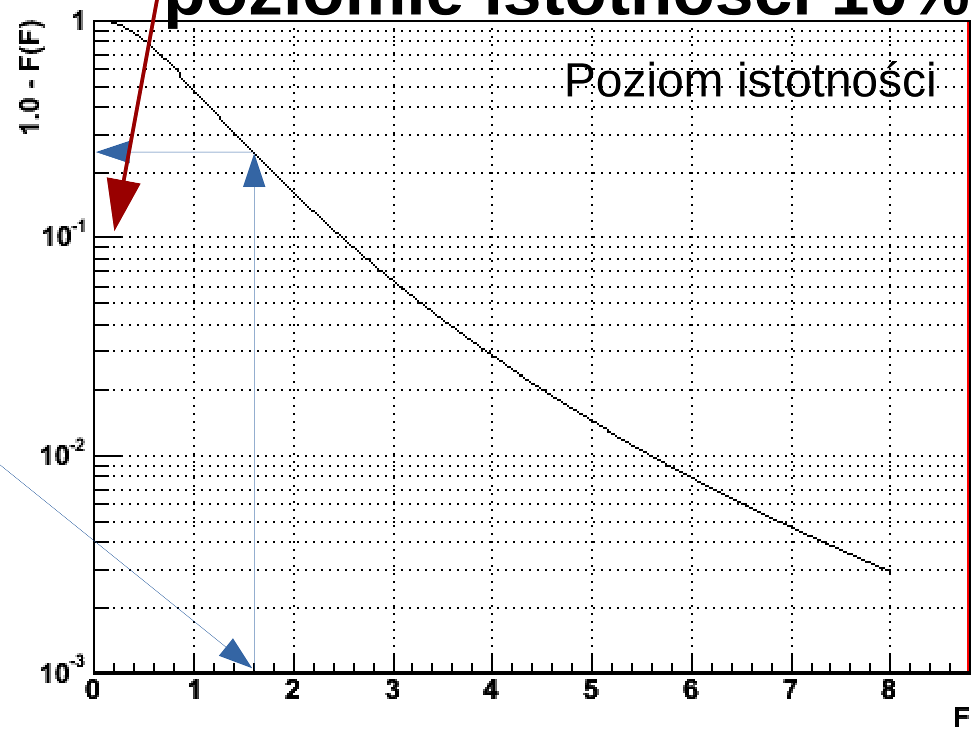
Numer pomiaru	Przyrząd 1	Przyrząd 2				
1	100	97	0,2	0,04	-2,86	8,16
2	101	101	1,2	1,44	1,14	1,31
3	102	102	2,2	4,84	2,14	4,59
4	100	99	0,2	0,04	-0,86	0,73
5	98	101	-1,8	3,24	1,14	1,31
6	97	98	-2,8	7,84	-1,86	3,45
7	100	101	0,2	0,04	1,14	1,31
8	101		1,2	1,44		
9	99		-0,8	0,64		
10	100		0,2	0,04		
Średnia	99,8	99,86				
Stopnie swobody	9	6				
S^2	19,6	20,86				
S^2/f	2,18	3,48				
F	1,6					

- Korzystamy z kwantyli funkcji F

$$F''_{0,2}(6,9)=F_{0,9}(6,9)=2.51$$
$$\boxed{F''_{0,1}(6,9)=F_{0,95}(6,9)=3.29}$$
$$F''_{0,02}(6,9)=F_{0,99}(6,9)=5.61$$
$$F''_{0,01}(6,9)=F_{0,995}(6,9)=6.89$$

- 1.6<3.29 – **nie ma podstaw do odrzucenia hipotezy na poziomie istotności 10%**





Porównanie wartości średnich (test t-Studenta)

Test równości średnich (t-Studenta)

- **Problem:** porównywanie wartości średnich dwóch prób losowych
- **Przykład:** badamy średni wzrost studentek 1 roku w Warszawie (populacja X) oraz w Nowym Jorku (populacja Y)
- **Pytanie (hipoteza):** czy wartości średnie obu populacji, na podstawie pobranych prób losowych, są jednakowe?
- Tak postawiona hipoteza cicho zakłada, że X i Y to te same populacje
- Powyższe rozważania możemy **uogólnić** na porównanie wartości średnich dwóch prób losowych z populacji X oraz Y o liczebnościach N_1 i N_2

Zastosowanie testu t-Studenta

- **Hipoteza:** równość wartości średnich z obu populacji: $\hat{x}=\hat{y}$
- Zakładamy (z centralnego twierdzenia granicznego), że wartości średnie z prób mają rozkład normalny z wariancjami średnich:

$$\sigma^2(\bar{X})=\sigma^2(X)/N_{1,}\quad \sigma^2(\bar{Y})=\sigma^2(Y)/N_2$$

- Wariancje średnich są estymowane przez estymatory:

$$s_{\bar{X}}^2=\frac{1}{N_1(N_1-1)}\sum_{j=1}^{N_1}(X-\bar{X})^2 \qquad s_{\bar{Y}}^2=\frac{1}{N_1(N_1-1)}\sum_{j=1}^{N_1}(Y-\bar{Y})^2$$

- Różnica wartości średnich z próby również ma rozkład zbliżony do normalnego: $\Delta=\bar{X}-\bar{Y}\Rightarrow\sigma^2(\Delta)=\sigma^2(\bar{X})+\sigma^2(\bar{Y})$
- Jeśli hipoteza jest prawdziwa, wówczas oczywiste jest, że $\hat{\Delta}=0$ oraz iloraz $\Delta/\sigma(\Delta)$ powinien podlegać rozkładowi Gaussa

Test różnic t-Studenta

- Skoro tak, to oczywiście $\sigma^2(X)=\sigma^2(Y)$, zatem można je estymować za pomocą jednego estymatora jako średnią ważoną z dwóch prób:

$$s^2=\frac{(N_1-1)s_X^2+(N_2-1)s_Y^2}{(N_1-1)+(N_2-1)}$$

- Wtedy możemy zdefiniować estymatory:

$$s_{\bar{X}}^2=s^2/N_{1,}\quad s_{\bar{Y}}^2=s^2/N_{2,}\quad s_{\Delta}^2=s_{\bar{X}}^2+s_{\bar{Y}}^2=\frac{N_1+N_2}{N_1\cdot N_2}s^2$$

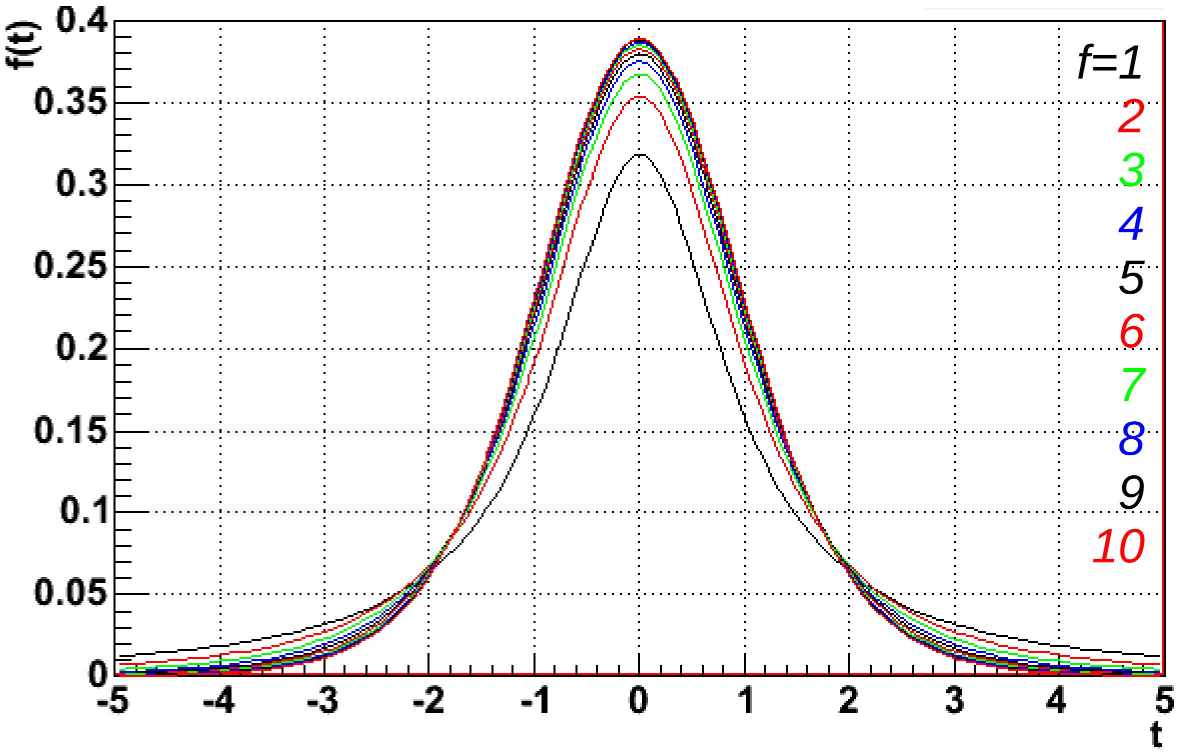
- Można udowodnić, że zmienna $\Delta/s(\Delta)$ podlega **rozkładowi t-Studenta** z liczbą stopni swobody $f=N_1+N_2-2$

- Równość wartości średnich można więc weryfikować posługując się **testem różnic Studenta**

- $\Delta/s(\Delta)$ obliczana jest na podstawie wyników dwóch prób. Jej wartość bezwzględną porównujemy z kwantylem rozkładu Studenta o liczbie stopni swobody f dla ustalonego poziomu istotności α. Sprawdzamy nierówność (**spełniona – odrzucamy hipotezę**):

$$|t|=\frac{|\Delta|}{s_\Delta}=\frac{|\bar{X}-\bar{Y}|}{S_\Delta}>t'_\alpha=t_{1-\frac{1}{2}\alpha}$$

Rozkład t-Studenta



- Jak widać, rozkład t-Studenta jest symetryczny względem 0
- Rozkład dąży do rozkładu Gaussa gdy $f \to \infty$
- Z symetrii względem 0 mamy związek (analogicznie jak dla Gaussa): $P(|t| \leq t) = 2F(|t|) - 1$

- Możemy wyznaczyć graniczne wartości $\pm t'_{\alpha}$ odpowiadające poziomowi istotności α poprzez całkę:

$$\int_0^{t'_{\alpha}} f(t)\,dt = \frac{1}{2}(1-\alpha), \quad gdzie \quad t'_{\alpha} = t_{1-\frac{1}{2}\alpha}$$

- Kwantyle $t'_{\alpha} = t_{1-\frac{1}{2}\alpha}$ są stablicowane dla różnych poziomów istotności α oraz liczby stopni swobody *f*
- Jest to dwustronny test t-Studenta (jednostronny analogicznie)

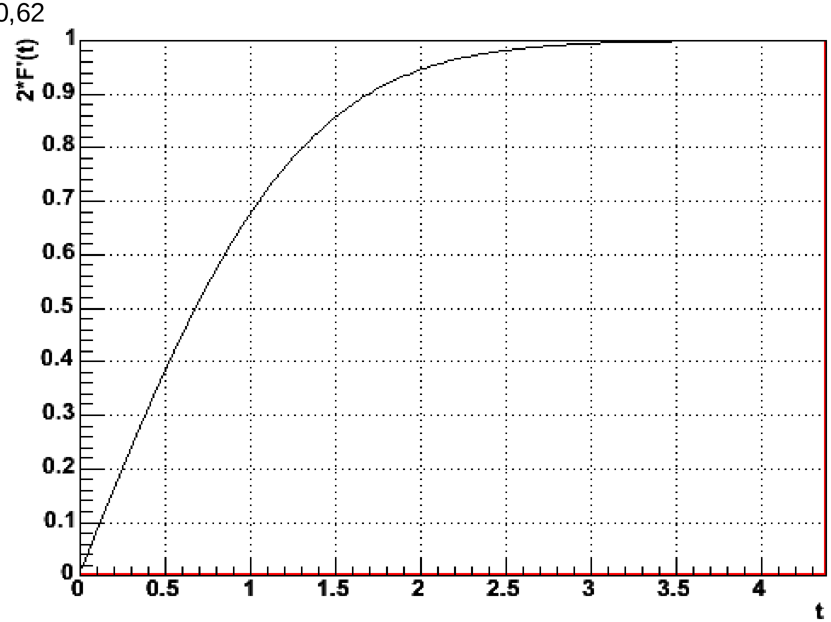
Test różnic t-Studenta - przykład

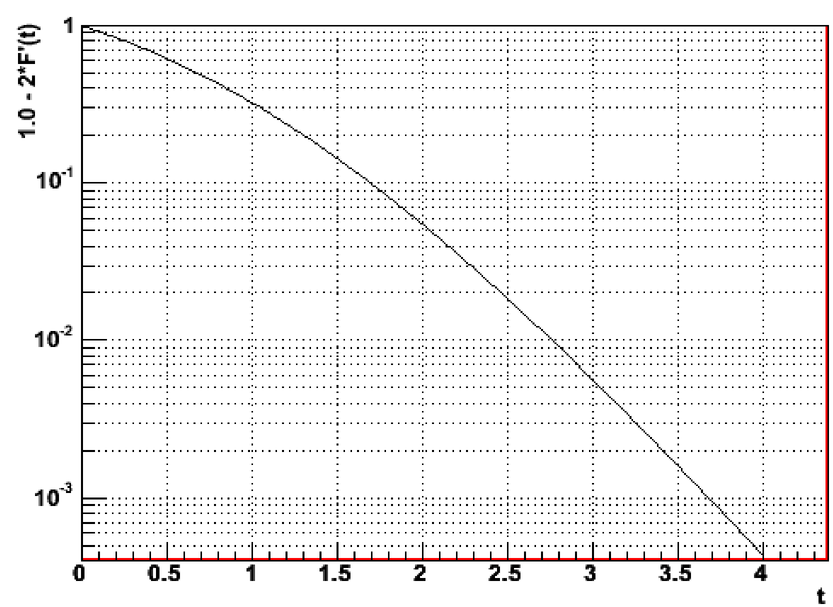
Numer pomiaru	Przyrząd 1	Przyrząd 2				
1	**100**	**97**	-0,21	0,04	-3,8	14,44
2	**101**	**101**	0,79	0,62	0,2	0,04
3	**102**	**102**	1,79	3,2	1,2	1,44
4	**100**	**99**	-0,21	0,04	-1,8	3,24
5	**98**	**101**	-2,21	4,89	0,2	0,04
6	**97**	**108**	-3,21	10,31	7,2	51,84
7	**100**	**101**	-0,21	0,04	0,2	0,04
8	**101**	**102**	0,79	0,62	1,2	1,44
9	**99**	**96**	-1,21	1,47	-4,8	23,04
10	**100**	**101**	-0,21	0,04	0,2	0,04
11	**98**		-2,21	4,89		
12	**101**		0,79	0,62		
13	**100**		-0,21	0,04		
14	**102**		1,79	3,2		
15	**103**		2,79	7,78		
16	**101**		0,79	0,62		
17	**99**		-1,21	1,47		
18	**100**		-0,21	0,04		
19	**102**		1,79	3,2		
Ilość pomiarów	19	10				
Średnia	**100,21**	**100,8**	-0,59			
Stopnie swobody	18	9				
S^2	43,16	95,6				
S^2/f	2,4	10,62				
S^2	49,1					
S^2 Delta	8,18					

- Mamy kwantyle:

$$t'_{0,2}(27)=t_{0,9}(27)=1.71$$
$$t'_{0,1}(27)=t_{0,95}(27)=2,05$$
$$t'_{0,02}(27)=t_{0,99}(27)=2,77$$
$$t'_{0,01}(27)=t_{0,995}(27)=3,05$$
$$t'_{0,004}(27)=t_{0,998}(27)=3,43$$
$$t'_{0,002}(27)=t_{0,999}(27)=3,69$$

- **Hipotezy nie można odrzucić**





Hipotezy zerowa i alternatywna

Błędy I i II rodzaju

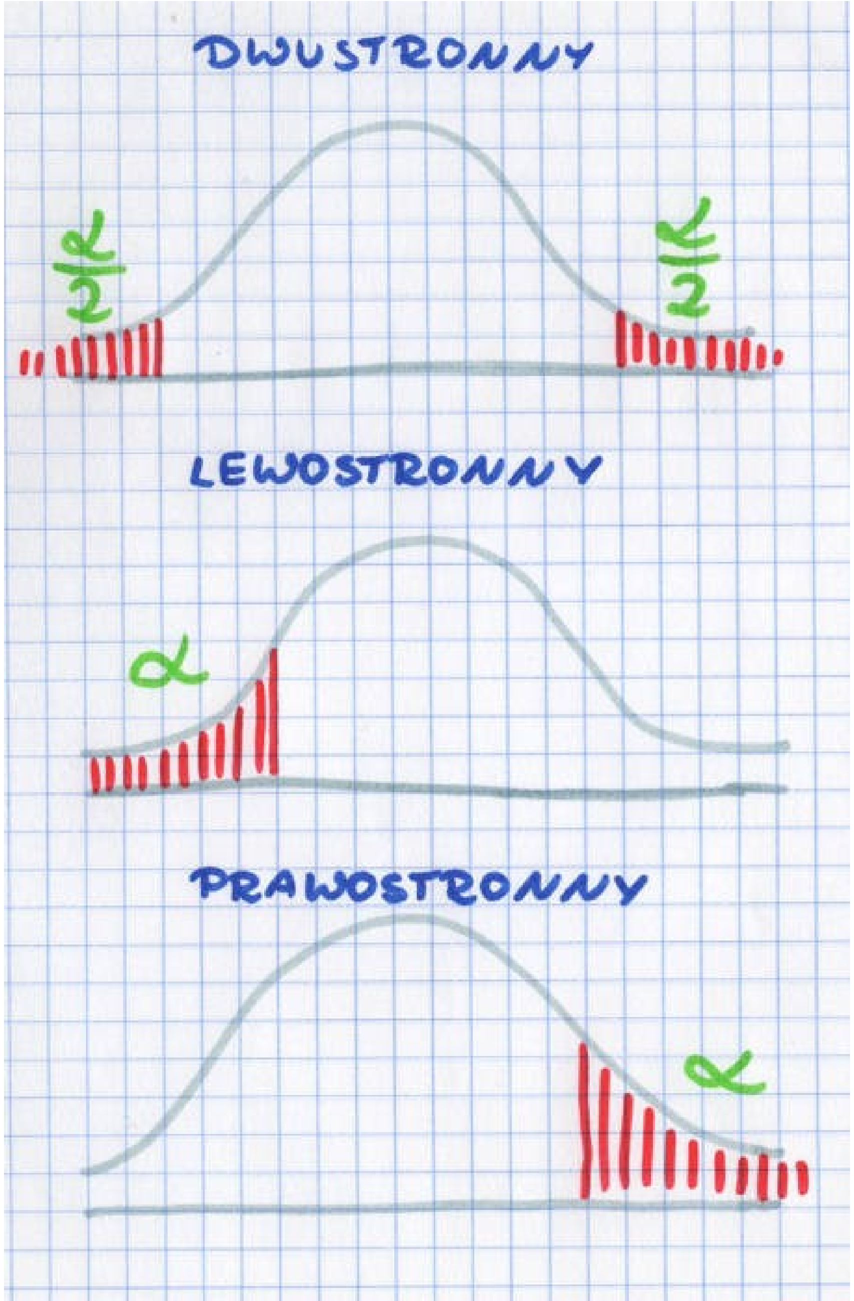
Na podstawie:
https://www.statystyczny.pl/hipotezy-statystyczne/

Hipoteza zerowa i alternatywna

- Hipoteza może być różna, przykłady hipotez:
  - średni wzrost Polaków to 175 cm
  - 2% dzieci w wieku szkolnym nie lubi czekolady
  - poziom szczęścia dwie minuty po zjedzeniu dużej porcji lodów jest wyższy niż przed zjedzeniem tejże dużej porcji lodów
- **Hipoteza zerowa** (H_0) to w uproszczeniu taka, gdy nie widzimy różnicy – np. czujemy się tak samo po zjedzeniu lodów jak przed
  - w przypadku pomiarów, np. wartość chi2 jest mała → teoria opisuje dane
- **Hipoteza alternatywa** (H_A) to przeciwieństwo hipotezy zerowej, którą <u>możemy zdefiniować na kilka sposobów</u>, np.:
  - np. jest różnica w szczęściu w zjedzeniu lodów (test dwustronny)
  - poziom szczęścia po zjedzeniu lodów jest mniejszy (test jednostronny)
  - poziom szczęścia po zjedzeniu lodów jest większy (test jednostronny)
  - w przypadku pomiarów chi2 jest duże → teoria nie opisuje danych (test jednostronny – rozkład chi2 jest niesymetryczny)
- **Statystyka testowa** - to funkcja próby, na podstawie której wnioskuje się o odrzuceniu lub nie hipotezy statystycznej – wielkość mająca swój rozkład prawd.

Statystyka testowa

- **Statystyka testowa** - to funkcja próby, na podstawie której wnioskuje się o odrzuceniu lub nie hipotezy statystycznej – wielkość mająca swój rozkład prawdopodobieństwa
- Z naszej próby losowej (eksperymentu) dostajemy jedną wartość – ona znajduje się gdzieś w tym rozkładzie
- Obszar krytyczny (obszar odrzuceń) jest zawsze na końcu rozkładu
  - jeśli hipoteza mówi, że coś jest różne – dwustronny
  - mniejsze lub większe – jednostronny
- **Statystyka testowa ma swój (różny) rozkład zarówno dla H_0 jak i H_A!!!**



Błąd I rodzaju

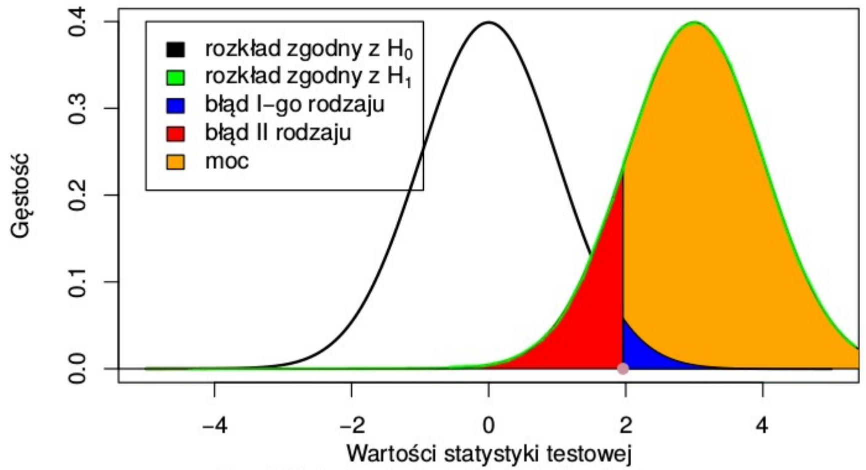
- **Błąd I rodzaju** to taki, gdzie odrzucamy hipotezę zerową a była ona prawdziwa
  - H_0: poziom szczęścia (średnio w populacji) po zjedzeniu dużej porcji lodów jest taki sam jak przed
  - H_A: poziom szczęścia (średnio w populacji) się zmienia
  - my na podstawie doświadczenia (próby losowej) zjedliśmy lody i np. przez te okropne wyrzuty sumienia, że znowu za dużo kalorii :) odrzucamy hipotezę zerową na rzecz alternatywnej
  - jeżeli w wyniku wielu prób losowych wynika, że jednak lądujemy w ogonie rozkładu owego szczęścia, to popełnimy właśnie błąd pierwszego rodzaju – bo odrzuciliśmy hipotezę zerową, która była prawdziwa
- Prawdopodobieństwo popełnienia błędu I rodzaju określa **poziom istotności** α, stąd oczekujemy, by był jak najmniejszy

Błąd II rodzaju

- **Błąd II rodzaju** to taki, gdy nie odrzucimy hipotezy fałszywej
  - ma on miejsce w sytuacji, kiedy jednak ten poziom szczęścia przed zjedzeniem lodów i po zjedzeniu się różni. Jeśli (**średnio w społeczeństwie**) te wyrzuty sumienia z powodu zjedzenia dobrych i smacznych lodów powodują, że poziom szczęścia zdecydowanie spada po zjedzeniu lodów, a my stwierdzimy (**na podstawie próby losowej**), że nie ma żadnej różnicy, to wtedy popełniamy błąd II rodzaju. **Nie odrzucamy hipotezy zerowej, mimo że jest ona fałszywa.**
- **Prawdopodobieństwo** popełnienia błędu II rodzaju określamy jako β

Moc testu

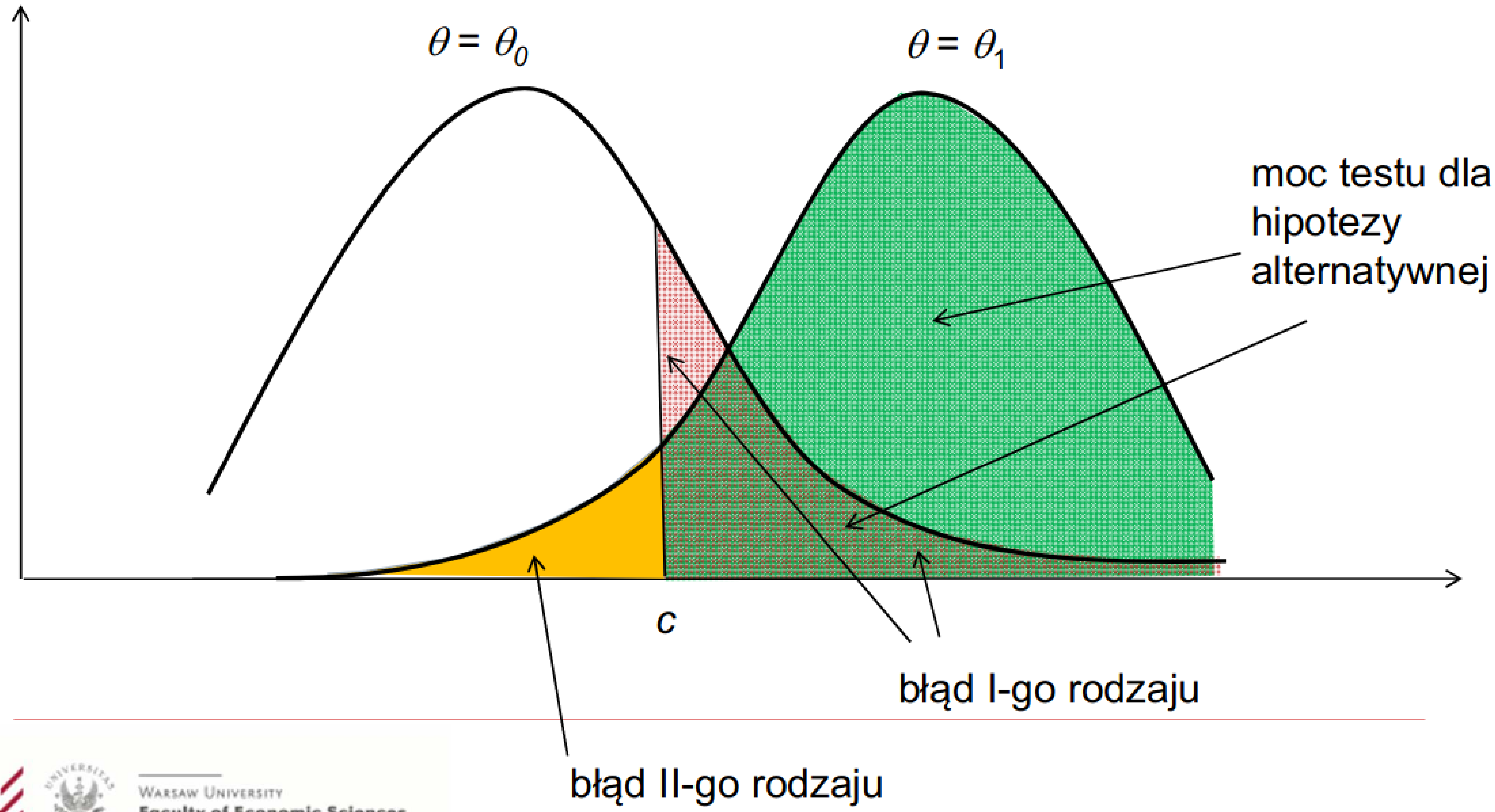
- Moc testu (prawdopodobieństwo, że prawidłowo odrzucimy hipotezę zerową) to 1-β. Inaczej mówiąc jest to prawdopodobieństwo niepopełnienia błędu II rodzaju.
- Moc testu zależy od kilku czynników:
  - Wielkości próby użytej w badaniu (im większa próba, tym większa moc testu).
  - Rzeczywistej wielkości efektu na tle losowej zmienności w populacji.
  - Przyjętego poziomu istotności α (między błędem I i II rodzaju jest taka zależność, że jeżeli zwiększamy prawdopodobieństwo popełnienia danego błędu, jednocześnie zmniejszamy je dla drugiego).



Rys.2 Relacje między błędem I, II rodzaju oraz mocą

Moc testu:
interpretacja graficzna (1)

rozkłady statystyki testowej przy założeniu prawdziwości hipotezy zerowej i alternatywnej



Warsaw University
Faculty of Economic Sciences

Testy statystyczne

KONIEC