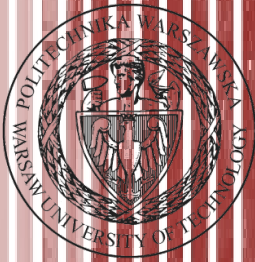


# Komputerowa analiza danych doświadczalnych

Wykład 9  
17.04.2019

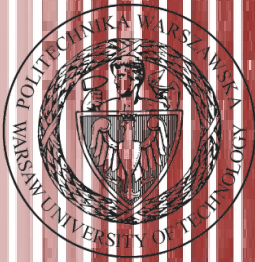
dr inż. Łukasz Graczykowski  
[lukasz.graczykowski@pw.edu.pl](mailto:lukasz.graczykowski@pw.edu.pl)

*Semestr letni 2018/2019*



# Metoda największej wiarygodności

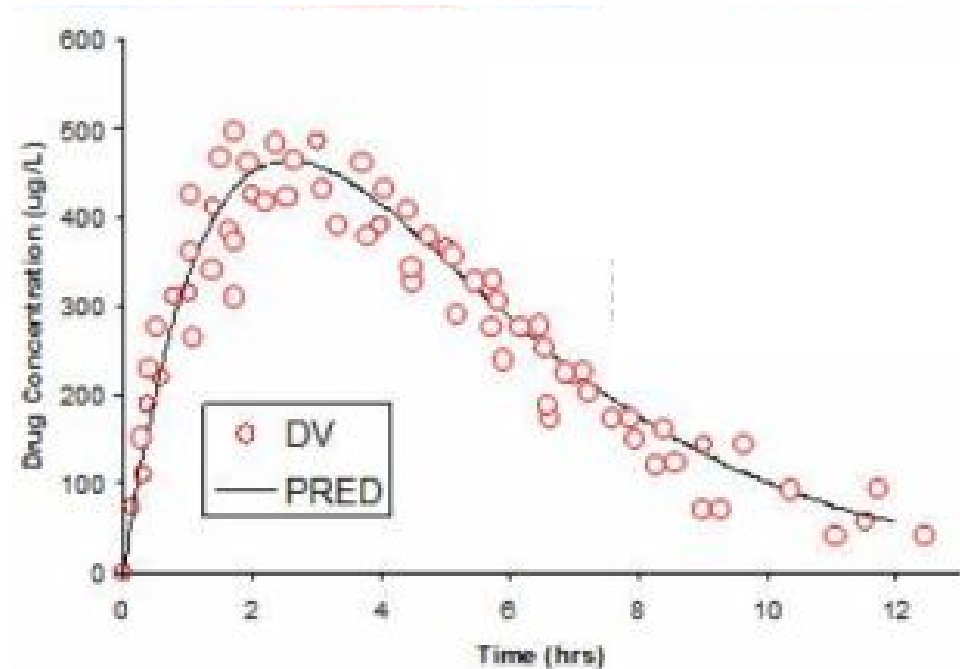
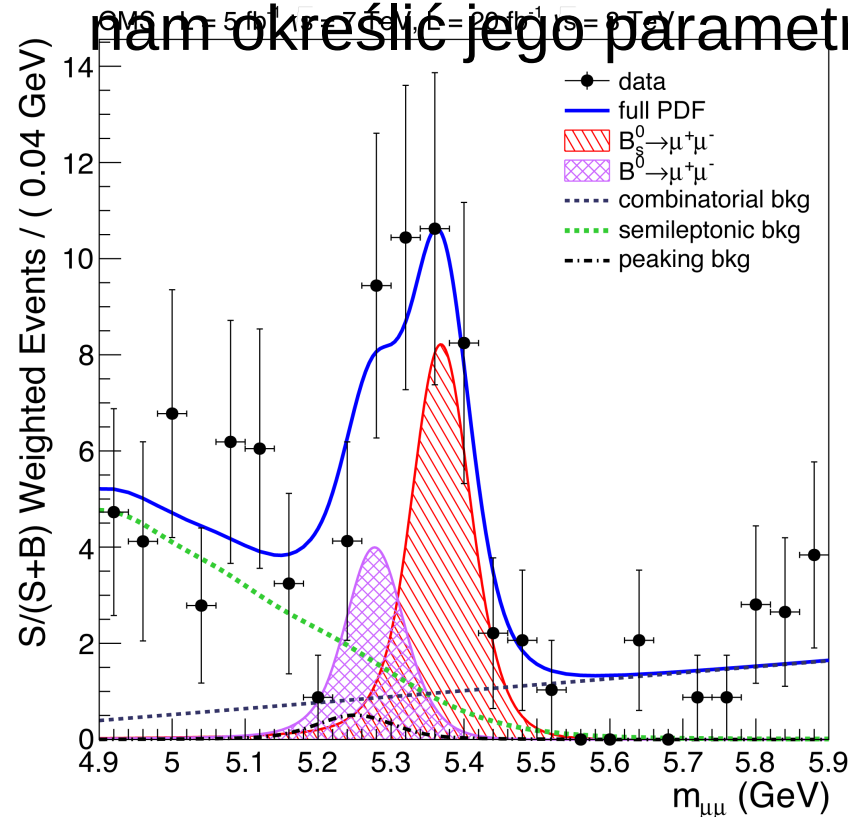
## Nierówność informacyjna



# Metoda największej wiarygodności

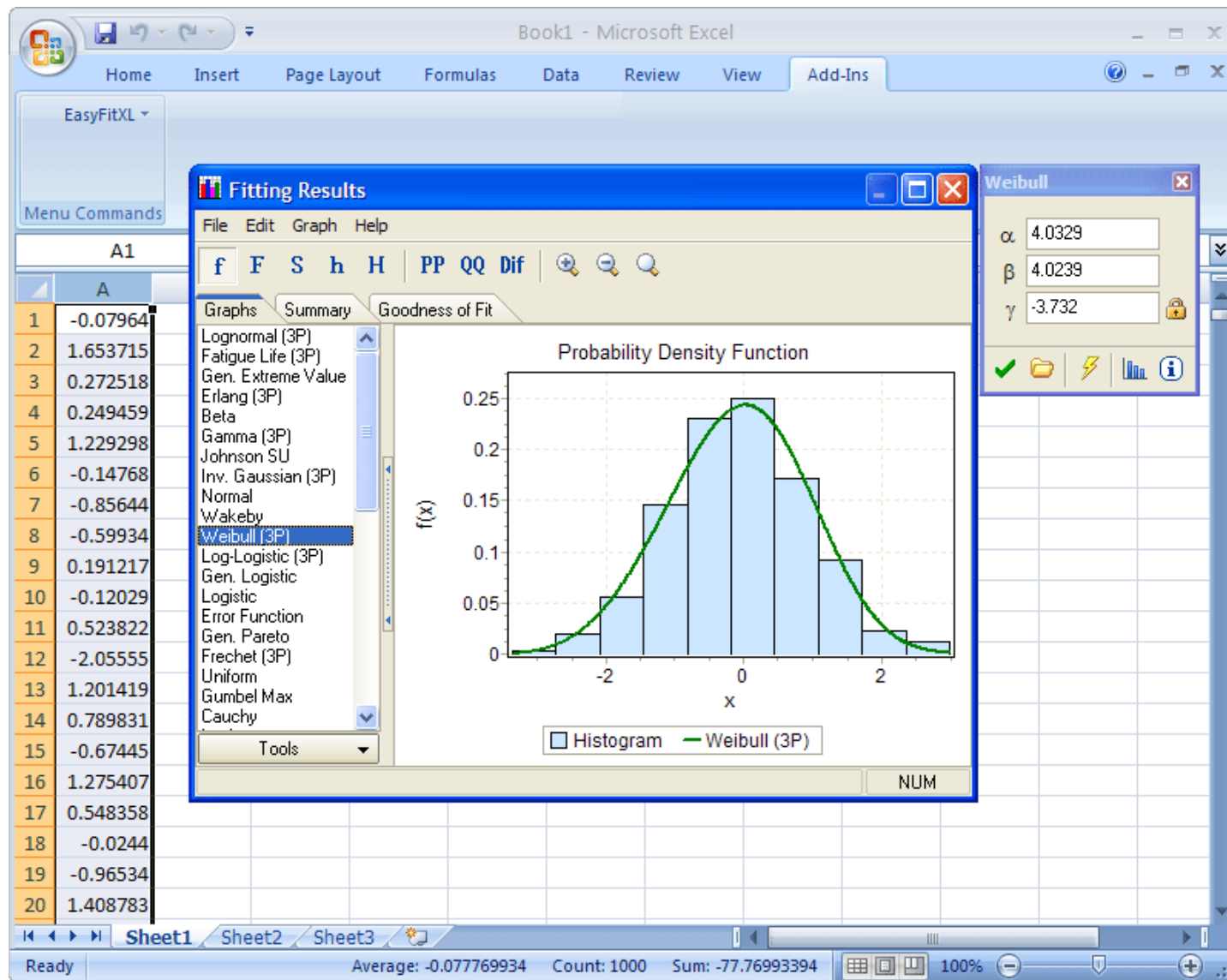
# “Magiczna” funkcja fit

- Do tej pory zajmowaliśmy się estymacją parametrów rozkładów, wprowadziliśmy estymatory i warunki, jakie powinny spełniać. Nie zajmowaliśmy się natomiast sposobem ich konstruowania (oprócz estymatora wartości oczekiwanej i wariancji)
- W praktyce mamy pewien zestaw danych (punkty pomiarowe) i chcemy do niego dopasować jakiś model teoretyczny, który pozwoli nam określić jego parametry





# “Magiczna” funkcja fit



<http://www.mathwave.com/articles/fit-distributions-excel.html>

# “Magiczna” funkcja fit

## 7.1 The Fit Method

The Fit method is implemented in ROOT for the histogram classes `TH1`, the sparse 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 xmin, Axis_t xmax)
```

- `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.

# Funkcja wiarygodności

- Do tej pory zajmowaliśmy się estymacją parametrów rozkładów, wprowadziliśmy estymatory i warunki, jakie powinny spełniać. Nie zajmowaliśmy się natomiast sposobem ich konstruowania (oprócz estymatora wartości oczekiwanej i wariancji)
- Problem ogólny:
  - mamy zbiór  $p$  parametrów:  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$
  - określony gęstością prawdopodobieństwa:  $f = f(\mathbf{x}; \lambda)$
  - dla  $n$  zmiennych losowych:  $\mathbf{X} = (X_1, X_2, \dots, X_n)$
- Jeden pomiar wielkości  $\mathbf{X}$ , czyli pobranie próby o liczności 1, prowadzi do uzyskania wyniku  $\mathbf{X}^{(j)} = (X_1 = x_1^{(j)}, X_2 = x_2^{(j)}, \dots, X_n = x_n^{(j)})$
- Takiemu doświadczeniu przypisujemy liczbę:
$$dP^{(j)} = f(\mathbf{x}^{(j)}; \lambda) d\mathbf{x}$$
- Która jest tzw. **prawdopodobieństwem a posteriori**. Mówi ona po uzyskaniu wyniku, jakie było prawdopodobieństwo jego uzyskania

# Funkcja wiarygodności

- **Prawdopodobieństwem a posteriori.** Mówi ona po uzyskaniu wyniku, jakie było prawdopodobieństwo jego uzyskania, czyli wartości  $\mathbf{x}^{(j)}$  takiej, że:  $x_i^{(j)} < \mathbf{x}^{(j)} \leq x_i^{(j)} + dx_i^{(j)}, (i=1,2,\dots,n)$

- Dla próby o  $N$  niezależnych elementach omawiane prawdopodobieństwo uzyskania wyniku  $\mathbf{x}^{(j)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}$  dane jest iloczynem:

$$dP = \prod_{j=1}^N f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}) d\mathbf{x}$$

- Gdy populacja jest charakteryzowana przez dwa różne zbiory parametrów,  $\lambda_1, \lambda_2$ , wtedy możemy określić **iloraz wiarygodności**:

$$Q = \frac{\prod_{j=1}^N f(\mathbf{x}^{(j)}; \lambda_1) d\mathbf{x}}{\prod_{j=1}^N f(\mathbf{x}^{(j)}; \lambda_2) d\mathbf{x}}$$

- Możemy go określić: “zbiór parametrów  $\lambda_1$  jest  $Q$  razy bardziej prawdopodobny niż zbiór parametrów  $\lambda_2$ ”



# Funkcja wiarygodności

- **Funkcją wiarygodności** nazywamy iloczyn postaci:

$$L = \prod_{j=1}^N f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

- Funkcja wiarygodności **jest zmienną losową** (jest funkcją próby)
- **Przykład:** iloraz wiarygodności – rzut asymetryczną monetą
  - na podstawie rzutu asymetryczną monetą i wyników tych rzutów chcemy ustalić, czy moneta należy do klasy A lub klasy B
  - założmy, że próba to 5 rzutów:  
1 orzeł i 4 reszki
  - stąd funkcja wiarygodności:

|        | A   | B   |
|--------|-----|-----|
| orzeł  | 1/3 | 2/3 |
| reszka | 2/3 | 1/3 |

$$L_A = \frac{1}{3} \cdot \left(\frac{2}{3}\right)^4 \quad L_B = \frac{2}{3} \cdot \left(\frac{1}{3}\right)^4 \quad Q = \frac{L_A}{L_B} = 8$$

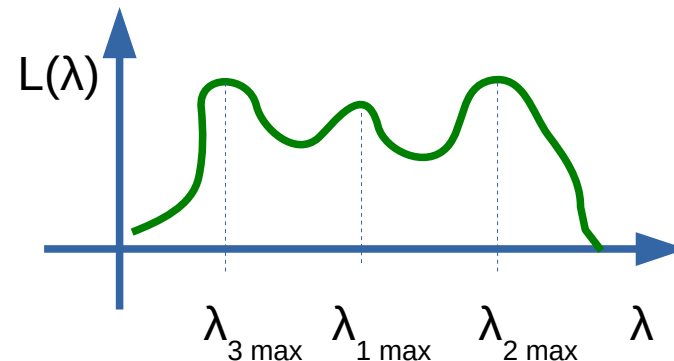
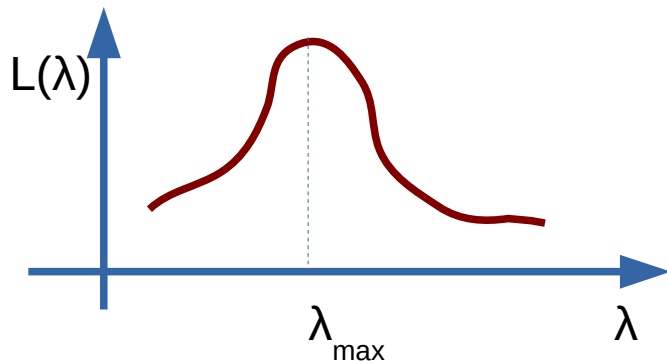
- z dużą dozą prawdopodobieństwa moneta należy do klasy A

# Metoda największej wiarygodności

- **Największą ufnością** obdarzamy zbiór parametrów, dla którego funkcja wiarygodności osiąga **maksymalną wartość**
- **Jak wyznaczyć maksimum?**
  - **warunek konieczny: przyrównać pierwszą pochodną  $L$  do zera**
- Różniczkowanie iloczynu jest jednak niewygodne, wprowadzamy więc **logarytm funkcji wiarygodności  $L$** :

$$l = \ln L = \sum_{j=1}^N \ln f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

- **Funkcją wiarygodności** jest odpowiednikiem gęstości prawdopodobieństwa, tylko określona dla parametrów. Ponieważ jest funkcją próby losowej, jest również zmienną losową



# Metoda największej wiarygodności

- W najprostszym przypadku wektor parametrów  $\lambda$  ma tylko jedną składową  $\lambda$
- Musimy wtedy rozwiązać **równanie wiarygodności**:  $l' = \frac{dl}{d\lambda} = 0$

- Czyli: 
$$l' = \sum_{j=1}^N \frac{d}{d\lambda} \ln f(\mathbf{x}^{(j)}; \lambda) = \sum_{j=1}^N \frac{f'}{f} = \sum_{j=1}^N \varphi(\mathbf{x}^{(j)}; \lambda)$$

- gdzie: 
$$\varphi(\mathbf{x}^{(j)}; \lambda) = \left( \frac{d}{d\lambda} f(\mathbf{x}^{(j)}; \lambda) \right) / (f(\mathbf{x}^{(j)}; \lambda))$$

jest **pochodną logarytmiczną**  $f$  względem  $\lambda$

- W przypadku gdy wektor  $\lambda$  ma  $p$  składowych, mamy układ  $p$  równań:  
$$\frac{\partial l}{\partial \lambda_i} = 0, \quad i = 1, 2, \dots, p$$

# Metoda największej wiarygodności – pods.

- Funkcją wiarygodności nazywamy iloczyn postaci:

$$L = \prod_{j=1}^N f(\mathbf{X}^{(j)}; \boldsymbol{\lambda})$$

- Funkcja wiarygodności **jest zmienną losową** (jest funkcją próby)
- **Największą ufnością** obdarzamy zbiór parametrów, dla którego funkcja wiarygodności osiąga **maksymalną wartość**
- **Jak wyznaczyć maksimum?**
  - warunek konieczny: przyrównać pierwszą pochodną  $L$  do zera
- Różniczkowanie iloczynu jest jednak niewygodne, wprowadzamy więc **logarytm funkcji wiarygodności**  $l$ :
$$l = \ln L = \sum_{j=1}^N \ln f(\mathbf{X}^{(j)}; \boldsymbol{\lambda}) = \sum_{j=1}^N \frac{f'(\mathbf{X}^{(j)}; \boldsymbol{\lambda})}{f(\mathbf{X}^{(j)}; \boldsymbol{\lambda})}$$
- **Funkcją wiarygodności** jest odpowiednikiem gęstości prawdopodobieństwa, tylko określona dla parametrów. Ponieważ jest funkcją próby losowej, jest również zmienną losową
- Równania wiarygodności:  $\frac{\partial l}{\partial \lambda_i} = 0, \quad i = 1, 2, \dots, p$

# Metoda największej wiarygodności

- **Przykład:** powtarzanie pomiarów o różnej dokładności
  - pomiar danej wielkości różnymi przyrządami pomiarowymi:
    - pomiary  $X^{(j)}$  będą się układały wokół wartości rzeczywistej  $\lambda$
    - niepewności będą miały rozkład normalny
    - mamy 1 wielkość, więc z wektora  $\mathbf{x}^{(j)}$  robi się jedna zmienna  $X^{(j)}$
  - **Wniosek:** pojedynczy pomiar to pobranie próby o liczebności 1 z rozkładu Gaussa o wartości średniej  $\lambda$  i wariancji  $\sigma_j$
  - Prawdopodobieństwo *a posteriori* możemy zatem wyrazić jako:

$$dP^{(j)} = f(X^{(j)}; \lambda) dx = \frac{1}{\sqrt{2\pi} \sigma_j} \exp\left(-\frac{(X^{(j)} - \lambda)^2}{2\sigma_j^2}\right) dx$$

- Zatem dla  $N$  pomiarów funkcja wiarygodności wynosi:

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



# Najbardziej wiarygodny wynik

- Logarytmiczna funkcja wiarygodności wynosi:

$$l = -\frac{1}{2} \sum_{j=1}^N \frac{(X^{(j)} - \lambda)^2}{\sigma_j^2} + \text{const}$$

- Równanie wiarygodności przybiera postać:

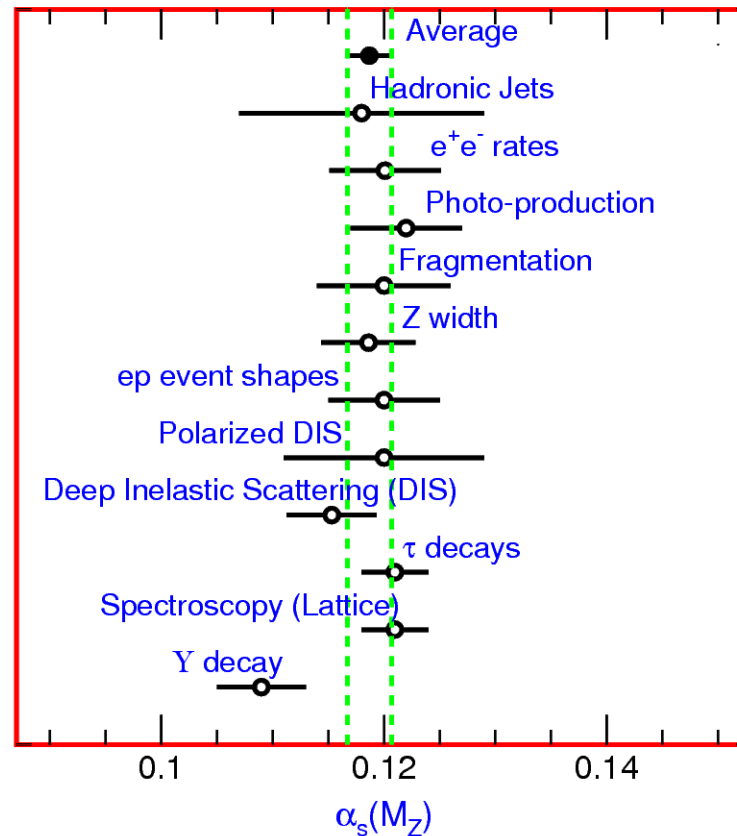
$$\frac{dl}{d\lambda} = \sum_{j=1}^N \frac{X^{(j)} - \lambda}{\sigma_j^2} = 0$$

- A jego rozwiązanie to:

$$\tilde{\lambda} = \sum_{j=1}^N \frac{X^{(j)}}{\sigma_j^2} / \sum_{j=1}^N \frac{1}{\sigma_j^2}$$

- **Wniosek:** wynik najbardziej wiarygodny jest średnią ważoną  $N$  pomiarów **tej samej cechy**, gdzie waga każdego pomiaru jest odwrotnością jego wariancji

# Uśrednianie wielu pomiarów – przykład



S. Eidelman et al.,  
Phys. Lett. B 592, 1 (2004)

**Figure 9.1:** Summary of the value of  $\alpha_s(M_Z)$  from various processes. The values shown indicate the process and the measured value of  $\alpha_s$  extrapolated to  $\mu = M_Z$ . The error shown is the *total* error including theoretical uncertainties. The average quoted in this report which comes from these measurements is also shown. See text for discussion of errors.

- Pomiar stałej sprzężenia oddziaływań silnych – szukanie najbardziej wiarygodnego wyniku → uśredniamy wszystkie wyniki z różnych pomiarów ważąc je odwrotnością wariancji

# Przykład – estymacja rozkładu norm.

- **Przykład:** badamy rozkład wagi studentek amerykańskich college'ów
  - rozkład wagi studentek w populacji jest opisany rozkładem normalnym o wartości średniej  $\mu$  i wariancji  $\sigma^2$
  - założmy, że wybraliśmy  $N=10$  reprezentantów, których wagi (w kg) układają się następująco:
    - 52,2 55,3 59,0 57,6 67,6 72,6 68,9 62,6 67,6 81,6
  - **Zadanie:** na podstawie wyniku pomiaru znajdź najbardziej wiarygodną estymację parametrów rozkładu
  - oczywiście funkcja rozkładu prawdopodobieństwa każdego wyniku (wagi studentek) dana jest wzorem:
$$f(x^{(j)}; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-(x^{(j)} - \mu)^2}{2\sigma^2}\right)$$
  - konstruujemy funkcję wiarygodności:

$$L(\hat{\mu}, \hat{\sigma}) = \sigma^{-n} (2\pi)^{-n/2} \exp\left[-\frac{1}{2\hat{\sigma}^2} \sum_{j=1}^N (x^{(j)} - \hat{\mu})^2\right]$$

# Przykład – estymacja rozkładu norm.

- oraz logarytmiczną funkcję wiarygodności:

$$l(\hat{\mu}, \hat{\sigma}) = \ln L(\hat{\mu}, \hat{\sigma}) = \ln \left( \hat{\sigma}^{-N} (2\pi)^{-n/2} \left[ -\frac{1}{2\hat{\sigma}^2} \sum_{j=1}^N (x^{(j)} - \hat{\mu})^2 \right] \right) = -\frac{1}{2\hat{\sigma}^2} \sum_{j=1}^N (x^{(j)} - \hat{\mu})^2 + \text{const}$$

- równania wiarygodności:

$$\frac{dl(\hat{\mu}, \hat{\sigma})}{d\hat{\mu}} = 0 \quad \frac{dl(\hat{\mu}, \hat{\sigma})}{d\hat{\sigma}} = 0$$

$$\frac{dl(\hat{\mu}, \hat{\sigma})}{d\hat{\mu}} = \frac{-2 \sum_{j=1}^N (x^{(j)} - \hat{\mu})(-1)}{2\hat{\sigma}} = 0 \Rightarrow \sum_{j=1}^N (x^{(j)}) - N\hat{\mu} = 0 \Rightarrow \hat{\mu} = \frac{\sum_{j=1}^N x^{(j)}}{n}$$

$$\frac{dl(\hat{\mu}, \hat{\sigma})}{d\hat{\sigma}} = -\frac{n}{2\hat{\sigma}} + \frac{\sum_{j=1}^N (x^{(j)} - \hat{\mu})}{2\hat{\sigma}^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{\sum_{j=1}^N (x^{(j)} - \hat{\mu})^2}{n}$$

- dla formalności powinniśmy jeszcze sprawdzić drugie pochodne...

# Przykład – estymacja rozkładu norm.

- liczymy estymatory:

$$\hat{\mu} = \frac{\sum_{j=1}^n x^{(j)}}{n} = 645/10 = 64,5$$

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^n (x^{(j)} - \hat{\mu})^2}{n} = 714/10 = 71.4$$

- otrzymaliśmy (poprzez metodę największej wiarygodności) estymację parametrów rozkładu normalnego całej populacji na podstawie pomiaru próby
- **Koniec**



# Przykłady funkcji wiarygodności



# Przykłady funkcji wiarygodności

- Rozkład Poissona

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

- Funkcja wiarygodności

$$l = \sum_{j=1}^N \{k^{(j)} \ln \lambda - \ln(k^{(j)}!) - \lambda\}$$

- oraz jej pochodna

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

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

- Średnia arytmetyczna z  $k$  jest nieobciążonym estymatorem  $\lambda$  o minimalnej wariancji równej  $\frac{\lambda}{N}$

- Rozkład dwumianowy

$$L(k; \lambda) = \binom{n}{k} \lambda^k (1-\lambda)^{n-k}$$

- Stąd:

$$l = \ln L = k \ln \lambda + (n-k) \ln (1-\lambda) + \ln \binom{n}{k}$$

$$l' = \frac{k}{\lambda} - \frac{n-k}{1-\lambda} = \frac{n}{\lambda(1-\lambda)} \left( \frac{k}{n} - \lambda \right)$$

- czyli średnia arytmetyczna  $k/n$  jest estymatorem o minimalnej wariancji równej  $\lambda(1-\lambda)/n$



# Nierówność informacyjna

# Nierówność informacyjna

- **Pytanie:** jak skonstruować estymator o optymalnych własnościach?
  - estymator jest **nieobciążony**, jeżeli **wartość obciążenia dla każdej próby:**  $B(\lambda) = E(S) - \lambda = 0$
  - oraz **(estymator zgodny) wariancja estymatora jest jak najmniejsza** (dąży do 0 dla liczebności próby losowej dążącej do nieskończoności):  $\sigma^2(S) - \text{minimalna}$
  - bardzo często istnieje jednak związek pomiędzy obciążeniem a wariancją i musimy szukać kompromisu – taki związek nazywamy **nierównością informacyjną**

- Po dłuższych przekształceniach (Brandt) otrzymujemy nierówność informacyjną – związek między obciążeniem, wariancją i informacją zawartą w próbce:

$$I(\lambda) = E(l'^2) = -E(l'') \quad l' = \sum_{j=1}^N \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)}$$

$$I(\lambda) = E(l'^2) = NE \left( \left( \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)} \right)^2 \right)$$

$$\frac{(B'(\lambda) + 1)^2}{I(\lambda)} \leq \sigma^2(S)$$

# Nierówność informacyjna

- **Nierówność informacyjna:**

$$\frac{(B'(\lambda)+1)^2}{I(\lambda)} \leq \sigma^2(S)$$

$$l' = \sum_{j=1}^N \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)}$$

- **Nierówność informacyjna** to związek pomiędzy obciążeniem, wariancją i informacją zawartą w próbie
- Jest to definicja ogólna. Prawa strona jest dolnym ograniczeniem dla wariancji danego (dowolnego) estymatora – ograniczenie to nazywamy **ograniczeniem minimalnej wariancji** albo **ograniczeniem Cramea-Rao**
- Kiedy obciążenie nie zależy od  $\lambda$ , a w szczególności gdy znika (równa się zero), nierówność redukuje się do:

$$\sigma^2(S) \geq \frac{1}{I(\lambda)}$$

# Nierówność informacyjna

- **Możemy teraz zadać pytanie:** jaki jest warunek osiągnięcia minimalnej wariancji? Innymi słowy, kiedy we wzorze Cramera-Rao zachodzi równość?

- Można pokazać (Brandt), że zajdzie tak, gdy:  $E(S) = B(\lambda) + \lambda$

$$l' = A(\lambda)(S - E(S)) \quad l = B(\lambda)S + C(\lambda) + D \quad L = d \exp(B(\lambda)S + C(\lambda))$$

- Wtedy w przypadku estymatora nieobciążonego o minimalnej wariancji otrzymujemy:

$$\sigma^2(S) = \frac{1}{I(\lambda)} = \frac{1}{E(l'^2)}$$

$$\sigma^2(S) = \frac{1}{|A(\lambda)|}$$

- Estymatory spełniające powyższe warunki nazywamy **estymatorami o najniższej wariancji**
- Jeżeli zamiast:  $L = d \exp(B(\lambda)S + C(\lambda))$  spełniony jest słabszy warunek:

$$L = g(S; \lambda) c(X^{(1)}, X^{(2)}, \dots, X^{(N)})$$

- Wtedy  $S$  nazywamy **estymatorem wystarczającym** dla  $\lambda$



**KONIEC**