Komputerowa analiza danych doświadczalnych

Wykład 10
29.04.2016

dr inż. Łukasz Graczykowski
lgraczyk@if.pw.edu.pl

Semestr letni 2015/2016

Metoda największej wiarygodności – przykład

Nierównosć informacyjna

Przykłady funkcji wiarygodności i prawo kombinacji niepewności

Własności asymptotyczne

Metoda największej wiarygodności (Maximum likelihood method)

Funkcja wiarygodności

- **Funkcją wiarygodności** nazywamy iloczyn postaci:

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

$$l = \ln L = \sum_{j=1}^{N} \ln f(\boldsymbol{X}^{(j)}; \boldsymbol{\lambda}) = \sum_{j=1}^{N} \frac{f'(\boldsymbol{X}^{(j)}; \boldsymbol{\lambda})}{f(\boldsymbol{X}^{(j)}; \boldsymbol{\lambda})}$$

- 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:

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

- **Funkcją wiarygodności** jest odpowiednikiem gęstości prawdopodobieństwa, tylko określona dla parametrów. Ponieważ jest funkcją próby losowej, jest również zmienną losową
- Równania wiarygodności: $\dfrac{\partial l}{\partial \lambda_i} = 0, \quad i = 1, 2, \ldots, p$

Przykład - estymacja param. rozkładu norm.

- **Przykład:** badamy rozkład wagi studentek amerykańskich college'ów
  - rozkład wagi studentek w populacji jest opisany rozkładem normalnym o wartości średniej μ i wariancji σ^2
  - załóżmy, że wybraliśmy N=10 reprezentantów, których wagi (w kg) $X^{(j)}$ 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) dana jest wzorem:

$$f\left(X^{(j)};\mu,\sigma\right)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-\left(X^{(j)}-\mu\right)^2}{2\sigma^2}\right)$$

  - konstruujemy funkcję wiarygodności:

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

Przykład - estymacja param. rozkładu norm.

- konstruujemy funkcję wiarygodności:

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

- oraz logarytmiczną funkcję wiarygodności:

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

- równania wiarygodności:

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

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

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

- dla formalności powinniśmy jeszcze sprawdzić drugie pochodne...

Przykład - estymacja param. rozkładu norm.

- liczymy estymatory:

$$\hat{\mu}=\frac{\sum_{j=1}^{n} X^{(j)}}{n}=645/10=64,5 \qquad \hat{\sigma}^2=\frac{\sum_{j=1}^{n}\left(X^{(j)}-\hat{\mu}\right)}{n}=714/10=71.4$$

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

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 dążacej do nieskończoności): $\sigma^2(S)$ - *minimalna*
  - bardzo często istnieje jednak związek pomiędzy obciążeniem a wiarancją i musimy szukać kompromisu – taki związek nazywamy **nierównością informacyjną**
- Oczywiste jest, że wariancja przybiera minimalną wartość, kiedy estymator jest **stałą** (wtedy wariancja wynosi po prostu 0)
- Rozpatrzmy estymator: $S(X^{(1)}, X^{(2)}, \ldots, X^{(N)})$
- Który jest funkcją próby: $\boldsymbol{X}=(X^{(1)}, X^{(2)}, \ldots, X^{(N)})$
- Łączna gęstość prawdopodobieństwa próby:

$$f(x^{(1)}, x^{(2)}, \ldots, x^{(N)}; \lambda)=f(x^{(1)}; \lambda) f(x^{(2)}; \lambda) \ldots f(x^{(N)}; \lambda)$$

Nierówność informacyjna

- Następnie po dłuższych przekształceniach **nierówność informacyjna**:

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

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

- Gdzie informacja próby ze wzgl. na λ:

$$I(\lambda) = E(l'^2) = -E(l'')$$

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

- **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 λ, 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 słówy, kiedy we wzorze Cramea-Rao zachodzi równość?
- Można pokazać, że zajdzie tak, gdy: $E(S)=B(\lambda)+\lambda$

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

- Wtedy w przypadku estymatora nieobciażonego o minimalnej wariancji otrzymujemy:

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

- Estymatory spełniające powyższe warunki nazywamy **estymatorami o najniższej wariancji**
- Jeżeli zamiast: $L=d\exp(B(\lambda)S+C(\lambda))$ spełniony jest słabszy warunek: $L=g(S;\lambda)c(X^{(1)},X^{(2)},\ldots,x^{(N)})$
- Wtedy S nazywamy **estymatorem wystarczającym** dla λ

Przykłady funkcji wiarygodności

Przykłady funkcji wiarygodności

- Rozkład Poissona

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

- Funkcja wiarygodności

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

- oraz jej pochodna

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

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

- Średnia arytmetyczna z k jest nieobciążonym estymatorem 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 $\lambda(1-\lambda)/n$

Prawo kombinacji niepewności

Prawo kombinacji niepewności

- Powróćmy do problemu z poprzedniego wykładu – wielokrotny pomiar tej samej wielkości z różną dokładnością, lub (równoważnie) problem pobierania próby z rozkładów normalnych o jednakowej wartości średniej λ i różnych, lecz znanych wariancjach σ_j
- Jak pamiętamy, logarytmiczna funkcja wiarygodności i je jochodna dla takiego przypadku:

$$l=-\frac{1}{2}\sum_{j=1}^{N}\frac{\left(X^{(j)}-\lambda\right)^2}{\sigma_j^2}+const \qquad l'=\frac{dl}{d\lambda}=\sum_{j=1}^{N}\frac{X^{(j)}-1}{\sigma_j^2}=0$$

- Możemy pochodną przekształcić:

$$l'=\sum_{j=1}^{N}\frac{X^{(j)}}{\sigma_j^2}-\sum_{j=1}^{N}\frac{\lambda}{\sigma_j^2}=\sum_{j=1}^{N}\frac{1}{\sigma_j^2}\left(\frac{\sum_{j=1}^{N}\frac{X^{(j)}}{\sigma_j^2}}{\sum_{j=1}^{N}\frac{1}{\sigma_j^2}}-1\right)$$

- I jego rozwiązanie: $S=\tilde{\lambda}=\sum_{j=1}^{N}\frac{X^{(j)}}{\sigma_j^2}/\sum_{j=1}^{N}\frac{1}{\sigma_j^2}$

- Estymator S jest nieobciążonym estymatorem wartości średniej λ.

Prawo kombinacji niepewności

- Jeśli teraz porównamy:

$$l' = \sum_{j=1}^{N} \frac{X^{(j)}}{\sigma_j^2} - \sum_{j=1}^{N} \frac{\lambda}{\sigma_j^2} = \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \left(\frac{\sum_{j=1}^{N} \frac{X^{(j)}}{\sigma_j^2}}{\sum_{j=1}^{N} \frac{1}{\sigma_j^2}} - \lambda \right) = \boxed{\sum_{j=1}^{N} \frac{1}{\sigma_j^2}} (\tilde{\lambda} - \lambda)$$

- Ze wzorem:

$$l' = A(\lambda)(S - E(S)) = A(\lambda)(S - B(\lambda) - \lambda)$$

$$dla \;\; S = \tilde{\lambda} \;\; i \;\; B(\lambda) = 0$$

$$l' = \boxed{A(\lambda)} (\tilde{\lambda} - \lambda)$$

- Widzimy, że estymator ten ma również minimalną wariancję. Jaka ona jest? Posłużmy się wzorem:

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

- Czyli: $\sigma^2(\tilde{\lambda}) = \dfrac{1}{\sum_{j=1}^{N} \frac{1}{\sigma_j^2}}$

- Wzór ten określamy jako **prawo kombinacji niepewności**, lub (bardziej kolokwialnie) **“uśrednianie niepewności w kwadratach”**

- Analogicznie jak w prawie dodawania niepewności

Prawo kombinacji niepewności

- Możemy zauważyć, że jeżeli wszystkie pomiary mają jednakową dokładność, otrzymamy:

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

- Analogicznie jak na Wykładzie 7, gdy liczyliśmy wartość oczekiwaną i wariancję z próby

Własności asymptotyczne estymatorów i funkcji wiarygodności

Własności asymptotyczne estymatorów i fun.

- Omówimy teraz niektóre własności funkcji wiarygodności i estymatorów największej wiarygodności dla dużych prób, czyli $N \to \infty$
- Estymator $S = \tilde{\lambda}$ jest zdefiniowany jako rozwiązanie równania wiarygodności:

$$l'(\lambda) = \sum_{j=1}^{N} \left(\frac{f'(X^{(j)};\lambda)}{f(X^{(j)};\lambda)} \right)_{\tilde{\lambda}} = 0$$

- Rozwijamy $l'(\lambda)$ na szereg Taylora wokół punktu $\lambda = \tilde{\lambda}$
- Czyli: $l'(\lambda) = l'(\tilde{\lambda}) + (\lambda - \tilde{\lambda}) l''(\tilde{\lambda}) + \ldots$
- Pierwszy wyraz z prawej strony znika (z uwagi na r. wiar.), drugą pochodną możemy zaś zapisac jako (w granicy$N \to \infty$)

$$l''(\tilde{\lambda}) = \sum_{j=1}^{N} \left(\frac{f'(X^{(j)};\lambda)}{f(X^{(j)};\lambda)} \right)'_{\tilde{\lambda}} \Rightarrow l''(\tilde{\lambda}) = NE\left(\left(\frac{f'(X^{(j)};\lambda)}{f(X^{(j)};\lambda)} \right)'_{\tilde{\lambda}} \right)$$

- Jeśli teraz posłużymy się wzorem: $I(\lambda) = E(l'^2) = -E(l'')$

Własności asymptotyczne estymatorów i fun.

- W takim przypadku możemy zapisać:

$$l''(\tilde{\lambda}) = E(l''(\tilde{\lambda})) = -E(l'^2(\tilde{\lambda})) = -I(\tilde{\lambda}) = -1/b^2 \leftarrow \text{stała}$$

analogicznie do

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

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

- **Innymi słowy:** zastąpiliśmy drugą pochodną logarytmicznej funkcji wiarygodności $l''(\tilde{\lambda})$, która jest funkcją próby, liczbą, która zależy jedynie od gęstości f i estymatora $\tilde{\lambda}$

- Czyli nasze rozwinięcie na szereg Taylora (pomijając wyższe wyrazy):

$$l'(\lambda) = (-1/b^2)(\lambda - \tilde{\lambda}) \Rightarrow l(\lambda) = (-1/2b^2)(\lambda - \tilde{\lambda})^2 + c$$

- Stałą c wzynaczamy podstawiając: $\lambda = \tilde{\lambda} \Rightarrow c = l(\tilde{\lambda})$

- I ostatecznie dostajemy (k=const):

$$\boxed{l(\lambda) - l(\tilde{\lambda}) = -\frac{1}{2}\frac{(\lambda - \tilde{\lambda})^2}{b^2} \Rightarrow L(\lambda) = k\exp\left(-\frac{(\lambda - \tilde{\lambda})^2}{2b^2}\right)}$$

- Funkcja wiarygodności (dla dużych N) ma więc **rozkład normalny**

Własności asymptotyczne estymatorów i fun.

- Funkcja wiarygodności (dla dużych N) ma więc **rozkład normalny**

$$\boxed{L(\lambda)=k\exp\left(-\frac{(\lambda-\tilde{\lambda})^2}{2b^2}\right)}$$

dla $\lambda=\tilde{\lambda},\ u(\tilde{\lambda})=b$ $\quad \lambda=\tilde{\lambda}\pm b$

otrzymujemy $-(l(\lambda)-l(\tilde{\lambda}))=\frac{1}{2}$

- Ponieważ estymowaliśmy parametr λ mamy więc: $S=\tilde{\lambda},\ E(S)=\lambda$
- Czyli $\tilde{\lambda}$ jest <u>nieobciążonym estymatorem o minimalnej wariancji wynoszącej</u>:

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

- Estymator ma wszystkie właściwości (nieobciążony, zgodny, z minimalną wariancją) **jedynie** w granicy $N\rightarrow\infty$
- Taki estymator nazywany jest estymatorem **asymptotycznie nieobciążonym**
- Analogicznie możemy powiedzieć, że funkcja wiarygodności jest **asymptotycznie normalna**

Własności asymptotyczne estymatorów i fun.

- Powiedzieliśmy sobie poprzednio, że funkcja wiarygodności jest miarą prawdopodobieństwa (rozkładem prawdopodobieństwa), ale dla <u>parametrów</u> badanego rozkładu prawd. danej cechy
- Zatem estymację parametru możemy przedstawić jako wynik i niepewność typu A:

$$\lambda=\tilde{\lambda},\ \ u(\lambda)=\sigma(\tilde{\lambda})$$

- Ponieważ funkcja wiarygodności asymptotycznie dąży do rozkładu normalnego, możemy powiedzieć, że wartość prawdziwa estymowanego parametru zawarta jest w przedziale:

$$\tilde{\lambda}-\sigma(\tilde{\lambda})<\lambda_0<\tilde{\lambda}+\sigma(\tilde{\lambda})$$

  z prawdopodobieństwem 68,3% („jeden sigma”)

- Podane związki możemy stosować jedynie dla prób o dużym N, dokładność przybliżeń zależy jednak również od rozpatrywanej funkcji gęstości – w każdym przypadku musimy zdecydować, czy powyższa procedura jest uprawniona do zastosowania

Przykład - wyznaczanie średniego czasu życia

- **Przykład:** chcemy zbadać średni czas τ życia jądra jakiegoś izotopu promieniotwórczego
- Prawdopodobieństwo że, istniejące w chwili t=0 jądro promieniotwórcze rozpadnie się w przedziale czasu od t do $t+dt$ wynosi:

$$f(t)\,dt = \frac{1}{\tau}\exp(-t/\tau)\,dt$$

- Badamy oczywiście pewną próbkę zawierającą wiele jąder – załóżmy, że zaobserwowaliśmy rozpady w chwilach: $t_{1,} t_{2,} \ldots, t_N$
- Konstruujemy zatem odpowiednią funkcję wiarygodności:

$$L(\tau) = \frac{1}{\tau^N}\exp\left(-\frac{1}{\tau}\sum_{i=1}^{N} t_i\right) = \frac{1}{\tau^N}\exp\left(-\frac{N}{\tau}\bar{t}\right)$$

- Obliczamy logarytm i pierwszą pochodną:

$$l(\tau) = \ln L(\tau) = -\frac{N}{\tau}\bar{t} - N\ln\tau \qquad l'(\tau) = \frac{N}{\tau}\left(\frac{\bar{t}}{\tau} - 1\right) = \frac{N}{\tau^2}(\bar{t} - \tau)$$

Przykład - wyznaczanie średniego czasu życia

- Porównując to ze wzorem:

$$l' = A(\lambda)(S - E(S)) = A(\lambda)(S - B(\lambda) - \lambda)$$

$$dla\ \ S = \tilde{\lambda}\ \ i\ \ B(\lambda) = 0$$

$$l' = A(\lambda)(\tilde{\lambda} - \lambda)$$

$$l'(\tau) = \frac{N}{\tau}\left(\frac{\bar{t}}{\tau} - 1\right) = \frac{N}{\tau^2}(\bar{t} - \tau)$$

- Widzimy, że rozwiązaniem o największej wiarygodności jest:

$$\tilde{\tau} = \bar{t} \qquad \sigma^2(\tilde{\tau}) = \tau^2 / N \Rightarrow dla\ \ \tau = \tilde{\tau} = t\ \ \Rightarrow \sigma^2(\tilde{\tau}) = \bar{t} / N$$

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

- Wstawiając $\tau = \tilde{\tau} = \bar{t}$ do wzoru na *l* otrzymamy:

$$l(\tilde{\tau}) = l_{max} = -N(1 + \ln \bar{t})$$

- Czyli:

$$-(l(\tau) - l(\tilde{\tau})) = N\left(\frac{\bar{t}}{\tau} + \ln\frac{\tau}{\bar{t}} - 1\right)$$

- W tym wyrażeniu trudno jest rozpoznać asymptotyczną postać równania:

$$l(\lambda) - l(\tilde{\lambda}) = -\frac{1}{2}\frac{(\lambda - \tilde{\lambda})^2}{b^2} \Rightarrow L(\lambda) = k\exp\left(-\frac{(\lambda - \tilde{\lambda})^2}{2b^2}\right)$$

Przykład - wyznaczanie średniego czasu życia

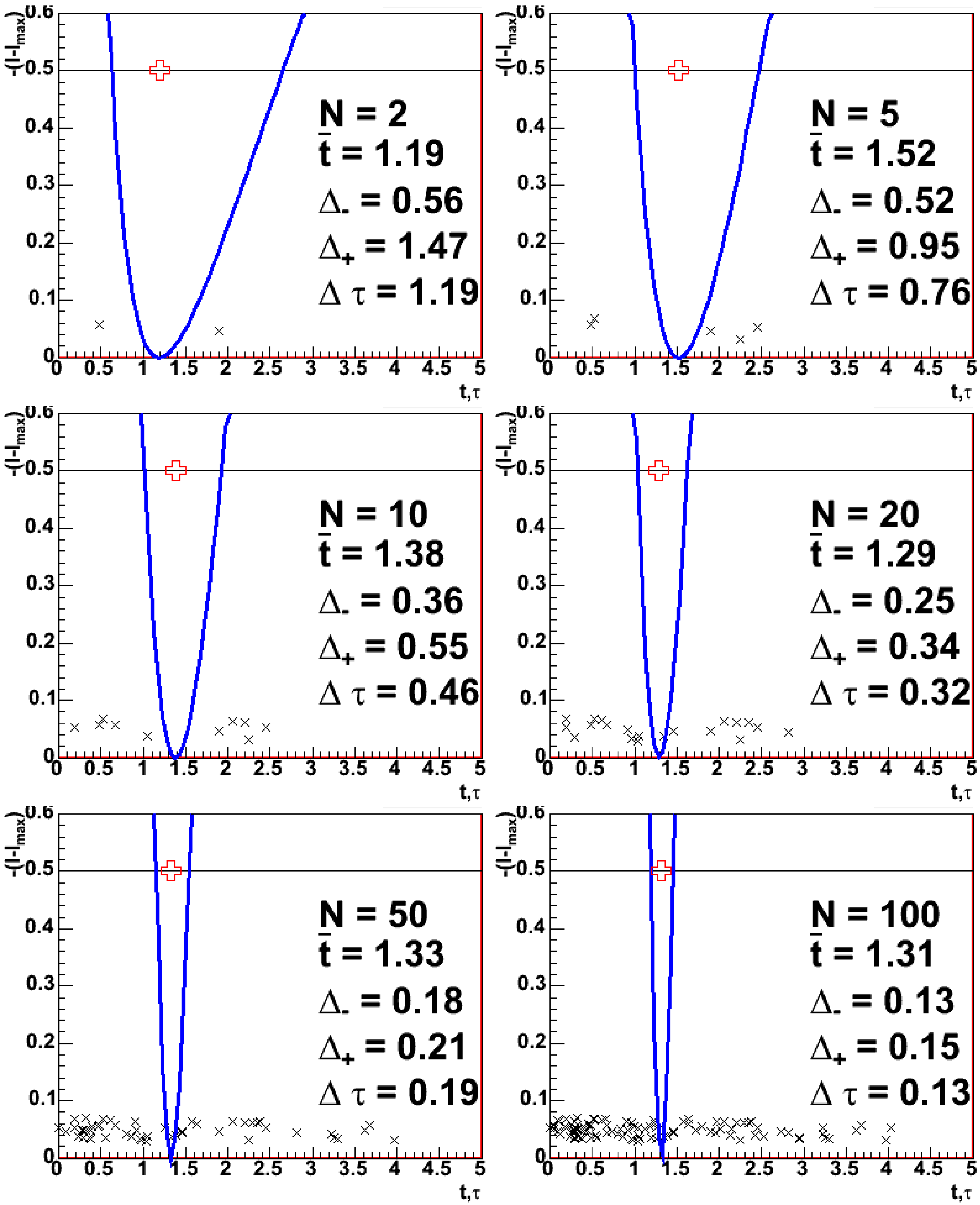
- Czyli nie możemy od razu zastosować rozumowania, że wartoć prawdziwa estymowanego mieści się w przedziale:

  $$\tilde{\lambda}-\sigma(\tilde{\lambda})<\lambda_0<\tilde{\lambda}+\sigma(\tilde{\lambda})$$

  z prawdopodobieństwem 68,3% ("1 sigma")
- Możemy jednak zdefiniować wartości τ_+ oraz τ_- takie, że:

  $$\tau_+=\tilde{\tau}+\Delta_+ \qquad \tau_-=\tilde{\tau}+\Delta_-$$
- Które spełniają warunek: $-\left(l(\tau_\pm)-l(\tilde{\tau})\right)=\frac{1}{2}$
- Wzór ten definiuje **niepewności niesymetryczne**
- W granicy $N\rightarrow\infty$ oczekiwalibyśmy: $\Delta_-,\Delta_+\rightarrow u(\tilde{\tau})=\sigma(\tilde{\tau})$
- Patrz rysunek na następnym slajdzie

Przykład - wyznaczanie średniego czasu życia



- Rysunek przedstawia funkcję:

$$-(l-l_{max})=-(l(\tau)-l(\tilde{\tau}))$$

- Punkty τ_+ , τ_- znajdują się na przecięciu z prostą:

$$-(l(\tau)-l(\tilde{\tau}))=1/2$$

- Krzyżem jest zaznaczony punkt: $\tilde{\tau}=\bar{t}$
- Widzimy, że wraz ze zwiększaniem się N funkcja

$$-(l-l_{max})=-(l(\tau)-l(\tilde{\tau}))$$

- Coraz bardziej przybiera kształst symetrycznej paraboli a niepewności asymetryczne coraz mniej się różnią od siebie

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:

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

- 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 $N \to \infty$ można zastąpić macierz A odpowiednimi wartościami oczekiwanymi $B = E(A)$
- Jeśli zaniedbamy wyrazy wyższych rzędów w rozwinięciu na szereg Taylora, dostaniemy wówczas funkcję wiarygodności postaci:

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

- Analogicznie jak dla jednego parametru, mamy tutaj p-wymiarowy rozkład normalny z macierzą kowariancji (dla estymatorów) $C = B^{-1}$
- Wariancje estymatorów największej wiarygodności $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_p$ dane są przez elementy diagonalne macierzy C: $\sigma^2(\tilde{\lambda}_i) = c_{ii}$
- Elementy pozadiagonalne są kowariancjami poszczególnych par estymatorów: $cov(\tilde{\lambda}_j, \tilde{\lambda}_k) = c_{jk}$
- 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%

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

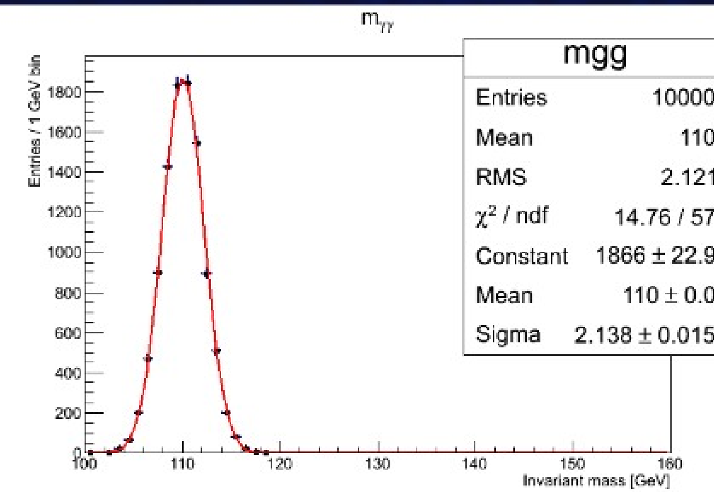
$$\mathcal{L} = \prod_{i=\text{bins}} \text{Poisson}\left[N_i^{\text{obs}}, A\exp\left(-\frac{1}{2}\frac{(x_i-\bar{x})^2}{\sigma^2}\right)\right]$$

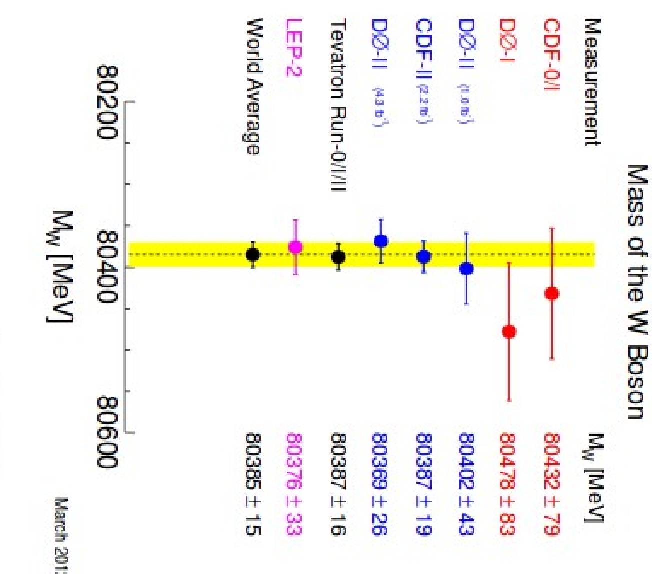
- **Unbinned data**
  - **Example:** normally distributed points about α

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

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

  - 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.

KONIEC