

Komputerowa analiza danych doświadczalnych

Wykład 9
21.04.2017

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

Semestr letni 2016/2017



Próby z odliczaniem, Próbki

Metoda największej wiarygodności

Nierówność informacyjna



Próba z odliczaniem, Próbki

Pobieranie próby z odliczaniem

- Często w praktyce pomiarowej mamy do czynienia z sytuacją, gdzie dokonujemy n obserwacji, z których tylko k ma pewną interesującą nas cechę. Resztę obserwacji, czyli $n-k$, odrzucamy
- Wybieramy (odliczamy) k z n elementów – jest to zatem rozkład dwumianowy (liczby sukcesów k) z prawdopodobieństwem p wystąpienia badanej cechy q – jej nie wystąpienia (nie znamy ich jednak → jest to przedmiot badań)
- **Pytanie:** jakie jest prawdopodobieństwo wystąpienia badanej cechy p ?
 - można udowodnić, że estymatorem p jest: $S(p) = \frac{k}{n}$
 - **Dlaczego?** rozkład dwumianowy to rozkład całkowitej liczby sukcesów, natomiast rozkład wielkości p jest rozkładem średniej liczby sukcesów
 - **Rozważmy przykład:** odbywają się wybory na Prezydenta RP, mamy dwóch kandydatów: BK i AD . Załóżmy, że p (60%) wszystkich wyborców preferuje AD . Jeśli jednak wybierzemy losowo 10 wyborców, mało prawdopodobne jest, że akurat 6 z nich wskaże AD (zależy to od tego “na kogo się trafi”) - **próbka** to realizacja próby los.

Pobieranie próby z odliczaniem

- **Rozważmy przykład:** odbywają się wybory na Prezydenta RP, mamy dwóch kandydatów: *BK* i *AD*. Załóżmy, że p (60%) wszystkich wyborców preferuje *AD*. Jeśli jednak wybierzemy losowo 10 wyborców, mało prawdopodobne jest, że 6 z nich wskaże *AD* (zależy to od tego “na kogo się trafi”)
- liczba wyborców *AD* w próbie losowej może wynieść trochę więcej lub trochę mniej niż 60%. Jeśli wielokrotnie powtórzymy tę czynność, otrzymamy rozkład (próbkiowania) wielkości p (*ang. sampling distribution*)
- widzimy, że rozkład p jest związany z rozkładem dwumianowym – rozkład dwumianowy to jednak rozkład całkowitej liczby sukcesów k (wyboru kandydata *AD*) a rozkład p to rozkład średniej liczby sukcesów
- średnia to oczywiście całość podzielona przez ilość, czyli rozmiar wybranej próby
- **Podsumowując:** rozkład dwumianowy – całkowita ilość (np. 6 z 10), rozkład p – średnia (0,6) → operują na innych argumentach

Pobieranie próby z odliczaniem

- Rozkład dwumianowy ma wartość oczekiwaną: np
- Wartość oczekiwana rozkładu p to zaś wartość oczekiwana rozkładu dwumianowego podzielona przez wielkość próby n : p
- Rozkład dwumianowy ma wariancję: $\sigma^2 = npq = np(1-p)$
- Czyli odchylenie standardowe (ma wymiar średniej): $\sigma = \sqrt{np(1-p)}$
- Dla rozkładu p musimy podzielić przez n , i wtedy otrzymamy odchylenie standardowe rozkładu p :
$$\sigma_p = \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}}$$
- Czyli wariancja rozkładu p : $\sigma_p^2 = \frac{p(1-p)}{n}$
- Jak już wspomnieliśmy, najlepszym estymatorem parametru p jest:
- Jaki jest zaś **estymator wariancji**? $S(p) = \frac{k}{n}$

- wstawiamy do wzoru na wariancję estymator wartości oczekiwanej, wtedy dostajemy:

$$s^2(S(p)) = \frac{1}{n} \frac{k}{n} \left(1 - \frac{k}{n}\right)$$
$$s(S(p)) = \frac{1}{n} \sqrt{k \left(1 - \frac{k}{n}\right)}$$

Niepewność statystyczna

- Zdefiniujmy teraz niepewność: $\Delta k = \sqrt{s^2(S(np))}$
- Wtedy otrzymamy: $\Delta k = \sqrt{k \left(1 - \frac{k}{n}\right)}$
- Niepewność ta zależy tylko od odliczonych elementów k oraz liczebności próby n i nazywana jest **niepewnością** (w starej nomenklaturze *błędem*) **statystyczną**
- Jeżeli k jest małe, czyli $k \ll n$, wtedy możemy wprowadzić parametr $\lambda = np$
$$W_k^n = \binom{n}{k} p^k q^{n-k} \quad \lim_{n \rightarrow \infty} W_k^n = f(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$
- W takim przypadku możemy (patrz Wykład 5) możemy uważać, że k jest pojedynczym elementem próby opisywanej rozkładem Poissona, wówczas otrzymujemy:
$$S(\lambda) = S(np) = k$$
$$\Delta \lambda = \sqrt{k}$$
- Czyli w **przybliżeniu** niepewność statystyczna liczby odliczeń k można zapisać jako łatwy do zapamiętania wzór: $\Delta k \approx \sqrt{k}$
- **Przykład:** niepewność zliczeń binu histogramu

$$S(np) = n \cdot \frac{k}{n} = k$$
$$\sigma_{np}^2 = np(1-p)$$
$$s^2(S(np)) = k \left(1 - \frac{k}{n}\right)$$

Dokładniejsza interpretacja niep. statyst.

- Jak pamiętamy, dla dużych k rozkład Poissona można przybliżyć rozkładem Gaussa o parametrach: $\mu = \lambda$, $\sigma^2 = \lambda$
- Dopóki liczba k nie jest zbyt mała (np. większa od 20, ale dużo mniejsza od $n!$) rozkład Poissona liczby k z parametrem λ możemy uważać za rozkład normalny zmiennej losowej X o powyższych parametrach → czyli zmienna skokowa zastąpiona jest zmienną ciągłą. Gęstość prawdopodobieństwa wynosi wtedy:

$$f(x; \lambda) = \frac{1}{\sqrt{\lambda} 2\pi} \exp\left(-\frac{(x-\lambda)^2}{2\lambda}\right)$$

- Za pomocą tej gęstości możemy zdefiniować granice **przedziału ufności** przy zadanym **poziomie ufności** $1-\alpha$

$$P(\lambda_m \leq \lambda \leq \lambda_p) = 1 - \alpha$$

- Prawdopodobieństwo wystąpienia prawdziwej wartości λ wewnątrz przedziału ograniczonego liczbami (λ_m, λ_p) jest równe poziomowi ufności $1-\alpha$

- Odpowiada to warunkom: $P(x > k | \lambda = \lambda_p) = 1 - \frac{\alpha}{2}$ $P(x < k | \lambda = \lambda_m) = 1 - \frac{\alpha}{2}$

Dokładniejsza interpretacja niep. statyst.

- Odpowiada to warunkom: $P(x > k | \lambda = \lambda_p) = 1 - \frac{\alpha}{2}$ $P(x < k | \lambda = \lambda_m) = 1 - \frac{\alpha}{2}$
- Można pokazać, że: $\frac{\alpha}{2} = \psi_0\left(\frac{k - \lambda_p}{\sigma}\right)$ $1 - \frac{\alpha}{2} = \psi_0\left(\frac{k - \lambda_m}{\sigma}\right)$

ψ_0 - dystrybuanta rozkładu normalnego

- Po dłuższych przekształceniach (Brandt) zakładając $\sigma^2 = k$ można pokazać, że na poziomie ufności $1 - \alpha = 68,3\%$ (czyli na poziomie 1σ), prawdziwa wartość k znajduje się w przedziale: $(k - \sqrt{k}, k + \sqrt{k})$

- Gdy **nie jest** spełniony warunek o dużym k (tj. nie przybliżamy Poissona rozkładem Gaussa), badamy rozkład:

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

- Można dowieść, że:

$$\frac{\alpha}{2} = F(k+1; \lambda_p) \quad 1 - \frac{\alpha}{2} = F(k; \lambda_m)$$

F - dystrybuanta rozkładu Poissona

$$F(k; \lambda) = \sum_{n=0}^{k-1} f(n; \lambda) = P(\mathbf{k} < k)$$

- Aby wyznaczyć λ_p, λ_m należy rozwiązać powyższe równania (numerycznie) - obliczyć funkcję odwrotną do rozkładu Poissona przy ustalonym k i zadanym prawdopodobieństwie (u nas $\frac{\alpha}{2}$ i $1 - \frac{\alpha}{2}$)

Pobieranie próbki w obecności tła

- W wielu doświadczeniach spotykamy się też z sytuacją, kiedy nie możemy określić, czy zaobserwowane zdarzenia są tego rodzaju jakie badamy (sygnał) lub innego typu (tło) – np. Badamy rozpad promieniotwórczy danego izotopu, ale mamy domieszkę innego typu, która rozpada się analogicznie
- Mamy więc wartość oczekiwaną liczby zdarzeń daną rozkładem Poissona w postaci sumy: $\lambda = \lambda_S + \lambda_T$
- Celem doświadczenia jest wyznaczenie λ_S
- Nie możemy skorzystać z poprzednich rozważań w celu wyznaczenia granic przedziału ufności

- Prawdopodobieństwo zaobserwowania n zdarzeń: $n = n_S + n_T$

$$f(n; \lambda_S + \lambda_T) = \frac{1}{n!} e^{-(\lambda_S + \lambda_T)} (\lambda_S + \lambda_T)^n$$

- Prawdopodobieństwa sygnału i tła: $f(n_S; \lambda_S) = \frac{1}{n_S!} e^{-\lambda_S} (\lambda_S)^{n_S}$

$$f(n_T; \lambda_T) = \frac{1}{n_T!} e^{-\lambda_T} (\lambda_T)^{n_T}$$

Pobieranie próbki w obecności tła

- Wiemy, że przy liczbie k zanotowanych przypadków, tło nie może tej liczby przewyższać. Wtedy wzór na tło zastępujemy wzorem:

$$f'(n_T; \lambda_T) = \frac{f(n_T; \lambda_T)}{\sum_{n_T=0}^k f(n_T; \lambda_T)}, \quad n_T \leq k$$

- W analogiczny sposób zastępujemy wzór zaobserwowania n zdarzeń:

$$f'(n; \lambda_S + \lambda_T) = \frac{f(n; \lambda_S + \lambda_T)}{\sum_{n_T=0}^k f(n_T; \lambda_T)}, \quad n_T \leq k$$

- Możemy teraz obliczyć granice przedziału $\lambda_{mS}, \lambda_{pS}$ ufności poszukiwanego parametru λ_S na poziomie ufności $1 - \alpha$:

$$\frac{\alpha}{2} = F'(k+1; \lambda_{pS} + \lambda_T)$$

$$1 - \frac{\alpha}{2} = F'(k; \lambda_{mS} + \lambda_T)$$

$$F'(k; \lambda_S + \lambda_T) = \sum_{n=0}^{k-1} f'(n; \lambda_S + \lambda_T) = P(\mathbf{k} < k)$$

F' - dystrybuanta rozkładu f'



Metoda największej wiarygodności

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: $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_p)$
 - określony gęstością prawdopodobieństwa: $f = f(\mathbf{x}; \boldsymbol{\lambda})$
 - 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:
$$dP^{(j)} = f(\mathbf{X}^{(j)}; \boldsymbol{\lambda}) d\mathbf{X}$$
- Która jest tzw. **prawdopodobieństwem a posteriori**. Mówi ona po uzyskaniu wyniku, jakie było prawdopodobieństwo jego uzyskania

Funkcja wiarygodności

- 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, 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:

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

- Gdy populacja jest ^{$j=1$} charakteryzowana przez dwa różne zbiory parametrów, λ_1, λ_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 λ_1 jest Q razy bardziej prawdopodobny niż zbiór parametrów λ_2 ”

Funkcja wiarygodności

- Takiemu doświadczeniu przypisujemy liczbę:

$$dP^{(j)} = f(\mathbf{X}^{(j)}; \lambda) dX$$

- **Funkcją wiarygodności** nazywamy iloczyn postaci:

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

- 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

	A	B
orzeł	1/3	2/3
reszka	2/3	1/3

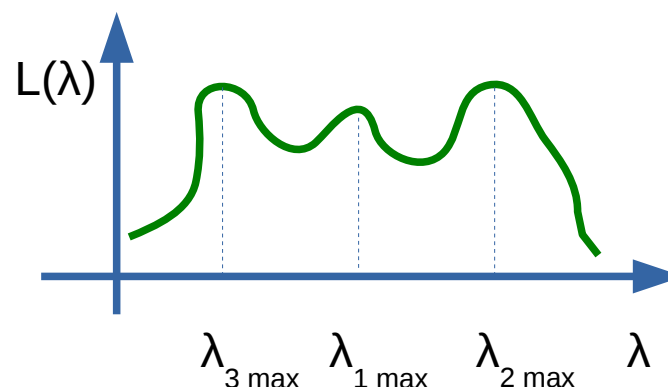
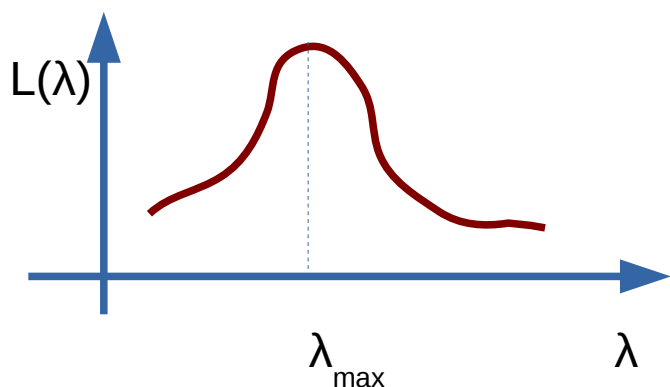
- stąd funkcja wiarygodności:

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

- 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** :
$$l = \ln L = \sum_{j=1}^N \ln f(x^{(j)}; \lambda)$$
- **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 λ ma tylko jedną składową λ
- Musimy wtedy rozwiązać **równanie wiarygodności**: $l' = \frac{dl}{d\lambda} = 0$
- Czyli: $l' = \sum_{j=1}^N \frac{d}{d\lambda} \ln f(\mathbf{x}^{(j)}; \lambda) = \sum_{j=1}^N \frac{f'}{f} = \sum_{j=1}^N \phi(\mathbf{x}^{(j)}; \lambda)$
- gdzie: $\phi(\mathbf{x}^{(j)}; \lambda) = \left(\frac{d}{d\lambda} f(\mathbf{x}^{(j)}; \lambda) \right) / (f(\mathbf{x}^{(j)}; \lambda))$
- jest **pochodną logarytmiczną** f względem λ
- W przypadku gdy wektor λ ma p składowych, mamy układ p równań:
$$\frac{\partial l}{\partial \lambda_i} = 0, \quad i = 1, 2, \dots, p$$
- **Przykład**: powtarzanie pomiarów o różnej dokładności
 - pomiar danej jednej wielkości różnymi przyrządami pomiarowymi:
 - pomiary $x^{(j)}$ będą się układały wokół wartości rzeczywistej λ
 - niepewności będą miały rozkład normalny
 - mamy 1 wielkość, więc z wektora $\mathbf{x}^{(j)}$ robi się jedna zmienna $x^{(j)}$

Metoda największej wiarygodności

- **Przykład:** powtarzanie pomiarów o różnej dokładności
 - pomiar danej wielkości różnymi przyrządami pomiarowymi:
 - pomiary $X^{(j)}$ będą się układały wokół wartości rzeczywistej λ
 - niepewności będą miały rozkład normalny
 - **Wniosek:** pojedynczy pomiar to pobranie próby o liczebności 1 z rozkładu Gaussa o wartości średniej λ i wariancji σ_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:

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

Metoda największej wiarygodności

- Logarytmiczna funkcja wiarygodności:

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

- Konstruujemy równanie wiarygodności:

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

- Którego 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, gdzie wagi są odwrotnościami wariancji poszczególnych pomiarów

Najbardziej wiarygodny wynik

- Logarytmiczna funkcja wiarygodności wynosi:

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

- Równanie wiarygodności przybiera postać:

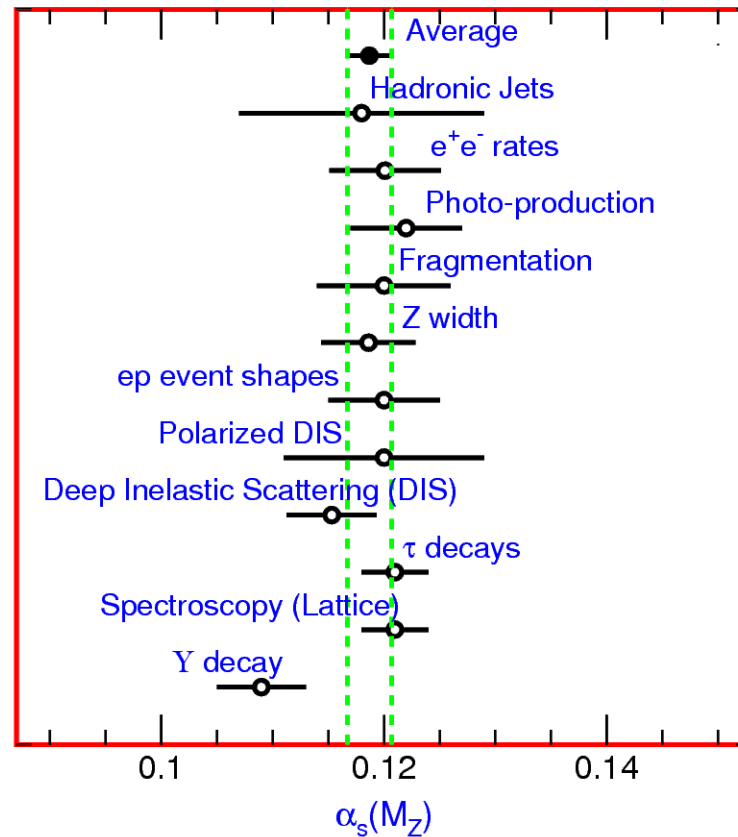
$$\frac{dl}{d\lambda} = \sum_{j=1}^N \frac{x^{(j)} - \lambda}{\sigma_j^2} = 0$$

- A jego rozwiązanie to:

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

- **Czyli** 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 α_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 z rozkładem dyskretnym

- Czasami nasze zjawisko opisywane jest rozkładem dyskretnym – estymowany parametr może przyjąć tylko dyskretne wartości
- **Przykład:** ze stawu, w którym pływa N ryb wyłowiono wpraw K ryb i po ich zaznaczeniu wpuszczono je z powrotem. W krótkim odstępie czasu wyłowiono n ryb, z których k było znaczonych.

Prawdopodobieństwo wynosi:

$$W_{n,k} = L(k; n, K, N) = \frac{\binom{N}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

- Znaleźć musimy taką wartość N , dla której funkcja L osiąga maksimum. Zbadajmy iloraz:

$$\frac{L(k; n, K, N)}{L(k; n, k, N-1)} = \frac{(N-n)(N-k)}{(N-n-K+k)N}$$

- Jest on mniejszy od 1, gdy $Nk > nK$ i większy od 1, gdy $Nk < nK$
- Czyli maksimum L osiągnie dla wart. całk. N najbliższej $N \approx nK/k$



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): $\sigma^2(S)$ - *minimalna*
 - bardzo często istnieje jednak związek pomiędzy obciążeniem a wariancją i musimy szukać kompromisu – taki związek nazywamy **nierównością informacyjną**
- Oczywiście jest, że wariancja przybiera minimalną wartość, kiedy estymator jest **stałą** (wtedy wariancja wynosi po prostu 0)
- Rozpatrzmy estymator: $S(X^{(1)}, X^{(2)}, \dots, X^{(N)})$
- Który jest funkcją próby: $\mathbf{X} = (X^{(1)}, X^{(2)}, \dots, X^{(N)})$
- Łączna gęstość prawdopodobieństwa próby:
$$f(x^{(1)}, x^{(2)}, \dots, x^{(N)}; \lambda) = f(x^{(1)}; \lambda) f(x^{(2)}; \lambda) \dots f(x^{(N)}; \lambda)$$

Nierówność informacyjna

- Wartość oczekiwana estymatora S wynosi więc:

$$E(S) = \int S(X^{(1)}, X^{(2)}, \dots, X^{(N)}) f(x^{(1)}; \lambda) f(x^{(2)}; \lambda) \dots f(x^{(N)}; \lambda) dx^{(1)} dx^{(2)} \dots dx^{(N)}$$

- Wyrażenie $E(S) = B(\lambda) + \lambda$ można zróżniczkować, po przekształceniu dostaniemy:

$$\begin{aligned} 1 + B'(\lambda) &= \int S \left(\sum_{j=1}^N \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)} \right) f(x^{(1)}; \lambda) \dots f(x^{(N)}; \lambda) dx^{(1)} dx^{(2)} \dots dx^{(N)} \\ &= E \left\{ S \sum_{j=1}^N \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)} \right\} = E \left\{ S \sum_{j=1}^N \phi(x^{(j)}; \lambda) \right\} = \boxed{E(S \cdot l')} \end{aligned}$$

- Następnie po dłuższych przekształceniach **nierówność**:

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

$$\begin{aligned} l' &= \sum_{j=1}^N \frac{d}{d\lambda} \ln f(x^{(j)}; \lambda) = \sum_{j=1}^N \frac{f'}{f} = \sum_{j=1}^N \phi(x^{(j)}; \lambda) \\ \phi(x^{(j)}; \lambda) &= \left(\frac{d}{d\lambda} f(x^{(j)}; \lambda) \right) / f(x^{(j)}; \lambda) \end{aligned}$$

- Gdzie informacja próby ze względu na λ : $I(\lambda) = E(l'^2) = NE \left(\left(\frac{f'(x; \lambda)}{f(x; \lambda)} \right)^2 \right)$

Nierówność informacyjna

- **Nierówność informacyjna:**

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

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

- **Nierówność informacyjna** to związek pomiędzy obciążeniem, wariancją i informacją zawartą w próbie
- Jest to definicja ogólna – nie wybraliśmy żadnego szczególnego estymatora. Prawa strona jest zatem 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:

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



KONIEC