

GRANICE I PRZEDZIAŁY UFNOŚCI

Roman Nowak

Klasyczne przedziały ufności

- konstrukcja neymanowska i inne podejścia klasyczne
- ich blaski i cienie

Bayesowskie przedziały wiarygodności (*credible intervals*)

Limit na masę cząstki Higgsa

Przedział ufności – wprowadzenie

Pomiar przekroju czynnego $\sigma = \mu/\mathcal{L}\varepsilon$, μ – oczekiwana liczba zdarzeń, ε – efektywność detekcji, \mathcal{L} – świetlność.

Poissonowski model z parametrem μ dla liczby k zliczeń.

Eksperyment $\rightarrow k = 3$.

Jakie wartości μ zgodzimy się zaakceptować?

$$P(k \leq 3 | \mu = 5) \approx 0,265 \quad P(k \geq 3 | \mu = 1,00) \approx 0,082$$

$$P(k \leq 3 | \mu = 10) \approx 0,010 \quad P(k \geq 3 | \mu = 0,50) \approx 0,014$$

$$P(k \leq 3 | \mu = 13) \approx 0,001 \quad P(k \geq 3 | \mu = 0,20) \approx 0,001$$

Konkluzja: przedział ufności $\mu \in [0,5; 10,0]$ na poziomie ufności $1 - \alpha = 1 - 0,010 - 0,014 = 0,976$

Jaką wartość α mam wybrać?

Przedział ufności (Neyman, 1937)

Recepta:

- zmienna losowa X o rozkładzie $f_X(x|\theta)$,
- trzy **zadane** liczby: $0 < \alpha < 1$ oraz $0 < \beta < \gamma < 1$, takie że:

$$\gamma - \beta = 1 - \alpha,$$

- przy ustalonej wartości θ , wynik x_0 pomiaru ustanawiamy kwantylami x_β i x_γ rzędu β i rzędu γ rozkładu $f(x|\theta)$:

$$P(X \leq x_\beta | \theta) = \beta, \quad P(X \leq x_\gamma | \theta) = \gamma$$

(kwantyl x_p rzędu p to taka wartość x_p zmiennej X , że $P(X \leq x_p) = p$)

- rozwiązania obu równań względem θ , wyznaczają **przedział ufności** $[\theta_{\min}, \theta_{\max}]$ na poziomie ufności $1 - \alpha$

$$1 - \alpha = P(x_\beta \leq X \leq x_\gamma | \theta) = P(\theta_{\min} \leq \theta \leq \theta_{\max} | x_0) \quad (???)$$

Konstrukcja Neymana

Przy ustalonej wartości θ , wyznaczamy kwantyle x_β i x_γ :

$$P(x \leq x_\beta | \theta) = \beta, \quad P(x \leq x_\gamma | \theta) = \gamma$$

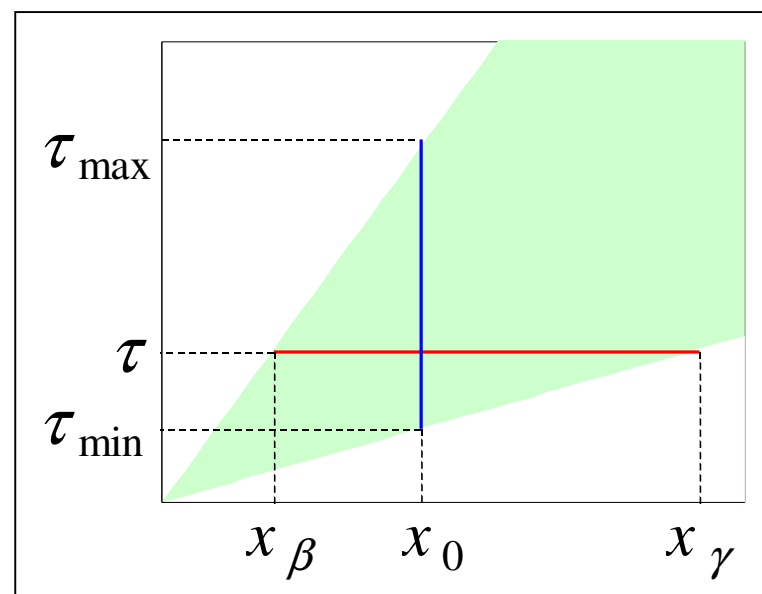
Przykład – rozkład wykładniczy:

$$P(x \leq x_\beta | \tau) = \frac{1}{\tau} \int_0^{x_\beta} \exp\left(-\frac{x}{\tau}\right) dx = \beta$$

$$x_\beta = -\tau \ln(1 - \beta)$$

$$P(x \leq x_\gamma | \tau) = \frac{1}{\tau} \int_0^{x_\gamma} \exp\left(-\frac{x}{\tau}\right) dx = \gamma$$

$$x_\gamma = -\tau \ln(1 - \gamma)$$



Wykres konstruujemy „w poziomie”, przedziały ufności odczytujemy „w pionie”.

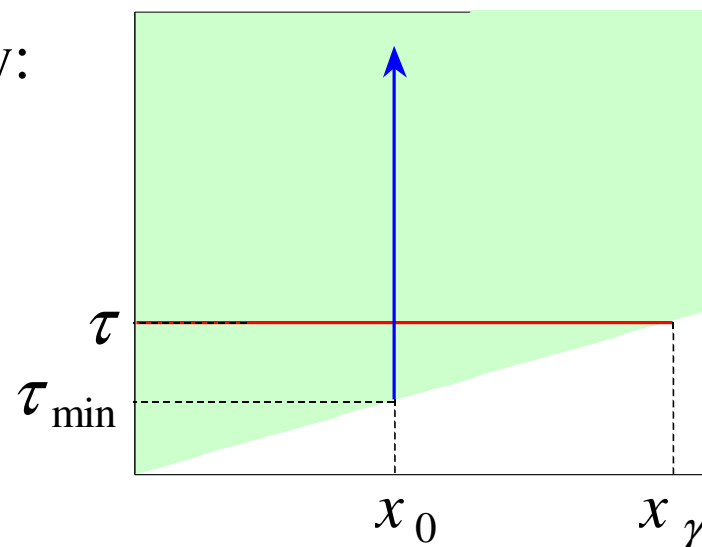
Dolna i górna granica

Dolna granica obszaru ufności, gdy:

$$x_{\beta} = „-\infty” \text{ (rząd } \beta = 0)$$

$$P(x \leq x_0 | \theta) = \gamma \Rightarrow \theta \in [\theta_{\min}, \infty]$$

jednostronny przedział ufności
na poziomie ufności $\gamma (= 1 - \alpha)$

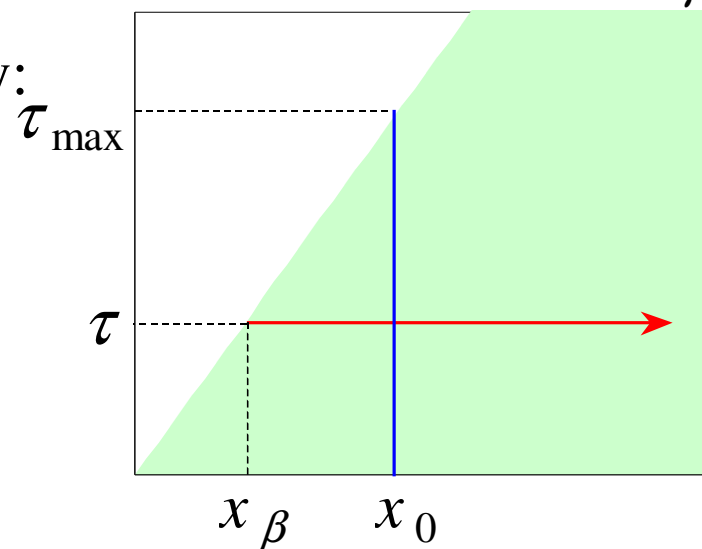


Górna granica obszaru ufności, gdy:

$$x_{\gamma} = „\infty” \text{ (rząd } \gamma = 1)$$

$$P(x \leq x_0 | \theta) = \beta \Rightarrow \theta \in [-\infty, \theta_{\max}]$$

jednostronny przedział ufności
na poziomie ufności $1 - \beta (= 1 - \alpha)$



Dolna i górna granica – przykład

Wyznaczanie efektywności detektora

Model dwumianowy: prawdopodobieństwo k sukcesów w n próbach. Pożądamy oceny parametru p :

$$\hat{p} \pm s_{\hat{p}} = \frac{k}{n} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}, \quad \text{ale:} \quad \hat{p} \pm s_{\hat{p}} = \begin{cases} 0, & k=0, \\ 1, & k=n, \end{cases} \quad \pm 0 !!!$$

Rozważny student:

$$\beta = 0 \Rightarrow P(k=0 | p) = (1-p)^n = 1-\alpha \Rightarrow p \in [0; 1 - \sqrt[n]{1-\alpha}]$$

$$\gamma = 1 \Rightarrow P(k < 100 | p) = p^n = 1-\alpha \Rightarrow p \in [\sqrt[n]{1-\alpha}; 1]$$

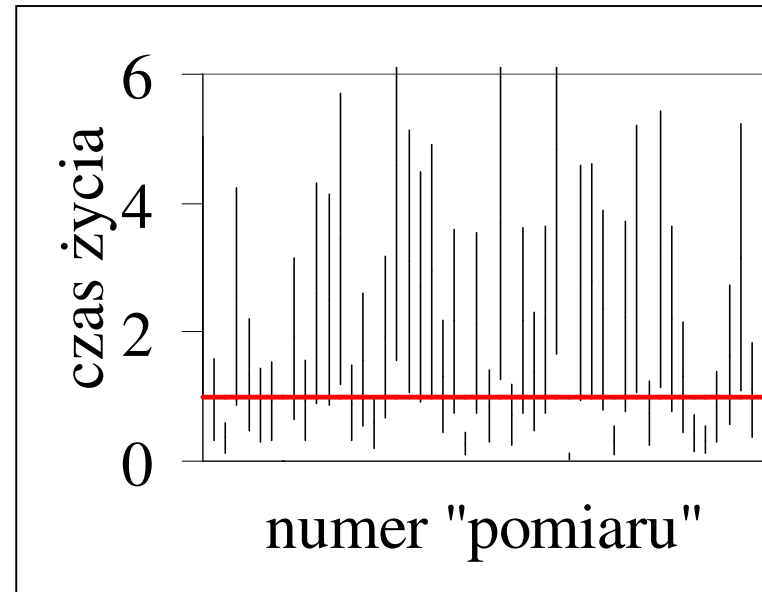
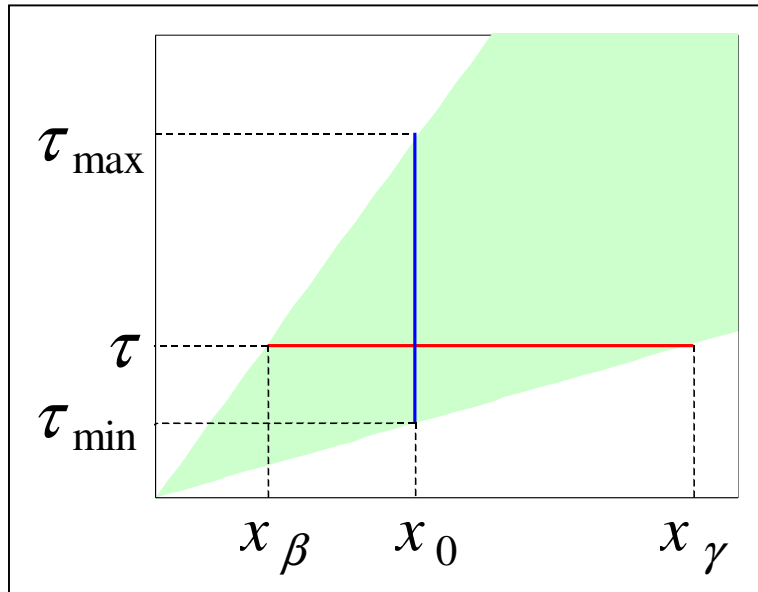
Dla $n = 100$ oraz $1 - \alpha = 0,95$: $p_{\max} \approx 0,03$, $p_{\min} \approx 0,97$

Jeszcze bardziej rozważny student: a może należałoby wyznaczyć granice także wtedy, gdy $k = 1, 2, 3, \dots$?

Przedział ufności – interpretacja

Zapis konwencjonalny:

$$P(x_\beta \leq x \leq x_\gamma | \theta) = P(\theta_{\min} \leq \theta \leq \theta_{\max} | x_0) = 1 - \alpha$$



Liczba $1 - \alpha$ określa prawdopodobieństwo:

- że przedział $[\theta_{\min}, \theta_{\max}]$, zawiera (***pokrywa***) parametr θ ,
a nie
- prawdopodobieństwo znalezienia parametru w tym przedziale!!!

Odkrycie (przez prasę) cząstki Higgsa

Dwie hipotezy:

- H_0 – istnieje cząstka Higgs z masą około 115 GeV
- H_1 – cząstka ta jest znacznie bardziej masywna lub nie istnieje

Wynik doświadczalny:

$P(\text{dane „jak są” lub obserwacja bardziej ekstremalna} | H_1) = 1\%$

Prasa: $P(H_1 | \text{dane}) = 1\%$, *ergo*: $P(H_0 | \text{dane}) = 99\%$

$$P(A|B) \neq P(B|A)$$

Przedział ufności – wariant I

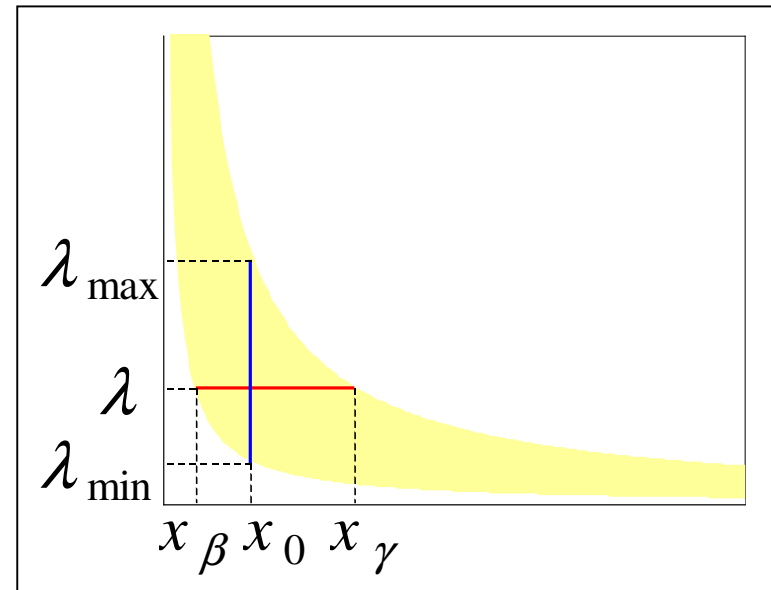
Skoro $\gamma - \beta = 1 - \alpha$, mamy jeden parametr swobodny

Konwencja **centralnego przedziału**

$$\beta = \frac{\alpha}{2}, \quad \gamma = 1 - \frac{\alpha}{2}$$

- zalecana (obok konwencji uporządkowanego stosunku wiarygodności) przez PDG
- zastosowania ograniczone do jednego parametru i jednej zmiennej
- przedziały niezmiennicze przy zamianie zmiennej i współzmiennicze przy reparametryzacji,

np. dla rozkładu wykładniczego, $1 - \alpha = 0,9$, $\tau = 1$, $x_0 = 1$:
przedział centralny: $\tau \in [0,33; 19,50]$, $\lambda \in [0,05; 3,0]$



Przedział ufności – wariant II

Konwencja **najkrótszego przedziału**

$$P(x \leq x_0 | \theta) = \gamma \Rightarrow \theta_{\max} = P^{-1}(x_0, 1 - \alpha + \beta),$$

$$P(x \leq x_0 | \theta) = \beta \Rightarrow \theta_{\min} = P^{-1}(x_0, \beta),$$

możemy więc poszukiwać przedziału o najkrótszej długości

$$\theta_{\max} - \theta_{\min} = \min(\beta)$$

- definiowalny w przypadku wielu parametrów
- trudne obliczenia
- granice zależą od parametryzacji

np. dla rozkładu wykładniczego, $1 - \alpha = 0,9$, $\tau = 1$, $x_0 = 1$:

przedział centralny: $\tau \in [0,33; 19,50]$, $\lambda \in [0,05; 3,0]$

przedział najkrótszy: $\tau \in [0; 9,5]$, $\lambda \in [0; 2,3]$

Przedział ufności – wariant III

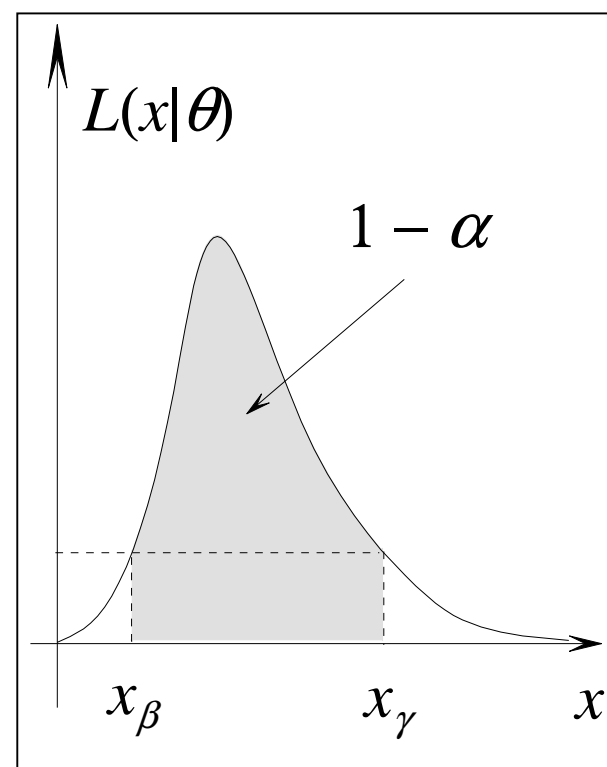
Konwencja **równej gęstości**

$$L(x_\beta | \theta) = L(x_\gamma | \theta)$$

$$\gamma - \beta = 1 - \alpha$$

$$L(x_\beta | \theta) = L(x_{1-\alpha+\beta} | \theta)$$

- zazwyczaj krótszy niż przedziały centralne
- definiowalny w przypadku wielowymiarowym
- trudne obliczenia
- granice zależą od zmiennej losowej



Przedział ufności – wariant IV

Konwencja **symetrycznego przedziału**:

$$\hat{\theta} - \theta_{\min} = \theta_{\max} - \hat{\theta} \quad \hat{\theta} \text{ ocena parametru } \theta$$

Atrakcyjny koncept, bo opisuje rodzaj symetrycznej niepewności

- ograniczenie do jednego parametru
- trudne obliczenia
- granice nie są współzmiennicze przy reparametryzacji

Przedział ufności – pokrycie

Absolutnie niezbywalna własność w podejściu klasycznym

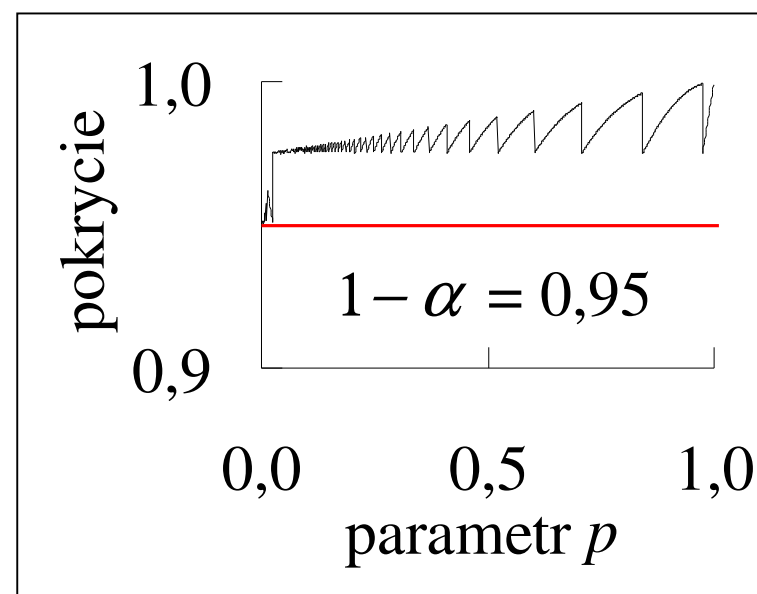
Nie jest to problem dla zmiennej ciągłej: $P(x_\beta \leq x \leq x_\gamma) = 1 - \alpha$

Zmienna dyskretna: $P_k(p) = p(1 - p)^{k-1}$, przedział centralny,
 $\alpha = 0,05$, $p = 0,5$:

$$k_{\alpha/2} = \frac{\ln(1 - \alpha/2)}{\ln(1 - p)} = 0,036$$

$$k_{1-\alpha/2} = \frac{\ln(\alpha/2)}{\ln(1 - p)} = 5,322$$

$$P\left(k_{\frac{\alpha}{2}} \leq k \leq k_{1-\frac{\alpha}{2}}\right) = \sum_{k=1}^6 P_k(p) = 0,984$$

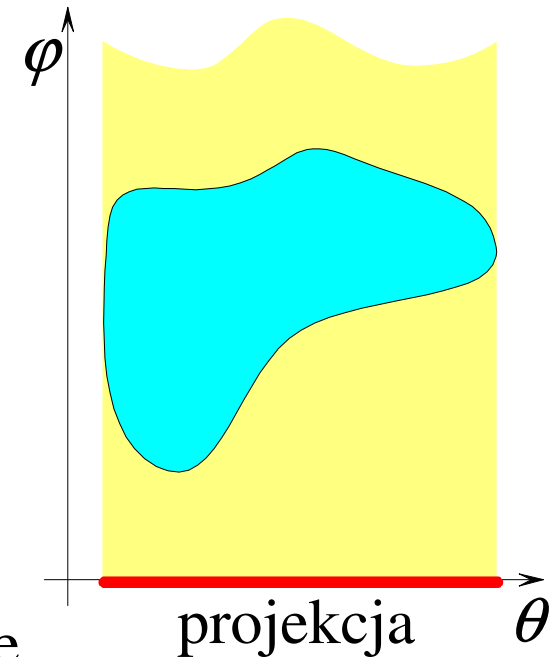


Zgoda na pokrycie nadmierne (konserwatywny przedział)

Parametr zakłócający (*nuisance*)

Statystyka x o rozkładzie $f(x|\theta, \varphi)$:

- θ – przedmiot zainteresowania
- φ – parametr uprzykrzający życie
- Formalne: wprowadzamy statystykę swobodną u dla parametru φ , a wtedy $f(x|\theta, \varphi) \rightarrow g(u|\theta)$: mało realistyczne (zazwyczaj utrata informacji)
- Faktoryzacja: $f(x|\theta, \varphi) = f_y(y|\theta) f_z(z|\varphi)$
- Zamiana parametru prowadząca do faktoryzacji
- Globalny przedział: dobry dla każdej wartości φ – trudne, nadmierne pokrycie
- Metoda często spotykana, ale błędna: ocenić φ z danych i podstawić – gwałci pokrycie!



Niefizyczne (puste) przedziały ufności

Pomiar masy antyneutrino elektronowego z rozpadu trytu:

$$m^2 = (-1,1 \pm 2,4) \text{ eV}^2$$

Centralny, gaussowski przedział ufności na poziomie 95%:

$$[\theta_{\min}, \theta_{\max}] = [x_0 - 1,96\sigma; x_0 + 1,96\sigma]$$

Gdyby $2,4 \text{ eV}^2 \rightarrow 0,6 \text{ eV}^2$, to $m^2 \in [-2,76; 0,08] \text{ eV}^2$.

Eksperymenty zliczające

s – liczba przypadków sygnału, o rozkładzie Poissona $\mathcal{P}(s|\mu)$

$$P(s \leq N | \mu) = \sum_{s=0}^N \mathcal{P}(s | \mu) = \alpha \Rightarrow \left\{ \begin{array}{l} \alpha = 0,1 \\ N = 0 \end{array} \right\} \Rightarrow \mu_{\max} = 2,30$$

b – liczba przypadków tła, o rozkładzie Poissona $\mathcal{P}(b|\lambda)$

$n = s + b$ – suma, podlega rozkładowi Poissona $\mathcal{P}(n|\mu + \lambda)$

$$P(n \leq N | \mu + \lambda) = \sum_{n=0}^N \mathcal{P}(n | \mu + \lambda) = \alpha$$

$N = 0, \alpha = 0,1$				
$\lambda =$	0	1	2	3
$\mu_{\max} =$	2,3	1,3	0,3	-0,7

Krótki kurs teorii testów

Własności testów:

- moc – prawdopodobieństwo odrzucenia fałszywej hipotezy
- zgodność – moc testu dąży do jedności ze wzrostem próbki
- brak obciążenia – szansa zaakceptowania fałszywej hipotezy H_0 jest mniejsza niż szansa jej odrzucenia gdy jest prawdziwa

Doskonałe narzędzie do produkcji przedziałów ufności!!!

Stosunek wiarygodności $R(x|\theta_0)$

$$R(x|\theta_0) = \frac{\mathcal{L}(x|\theta_0)}{\mathcal{L}(x|\hat{\theta}_{MW})} \leq 1$$

statystyka jednostajnie najmocniejszego, zgodnego i nieobciążonego testu hipotezy $H_0(\theta = \theta_0)$ przeciw $H_1(\theta \neq \theta_0)$.

Przedział ufności – wariant V

Konwencja **selektywnego przedziału**

Minimalizuje **ryzyko błędu drugiego rodzaju** – z minimalnym prawdopodobieństwem (na pewno mniejszym niż $1 - \alpha$) zawiera nieprawdziwą wartości parametru θ

- potężny i atrakcyjny koncept, ale niełatwy do wprowadzenia w życie, jeśli nie istnieje statystyka dostateczna
- przedział niezmienniczy względem zmiennej i współzmienniczy przy reparametryzacji

Konwencja **uporządkowanego stosunku funkcji wiarogodności** (*likelihood ratio ordering*), zwana **podejściem zunifikowanym**, daje przedziały selektywne

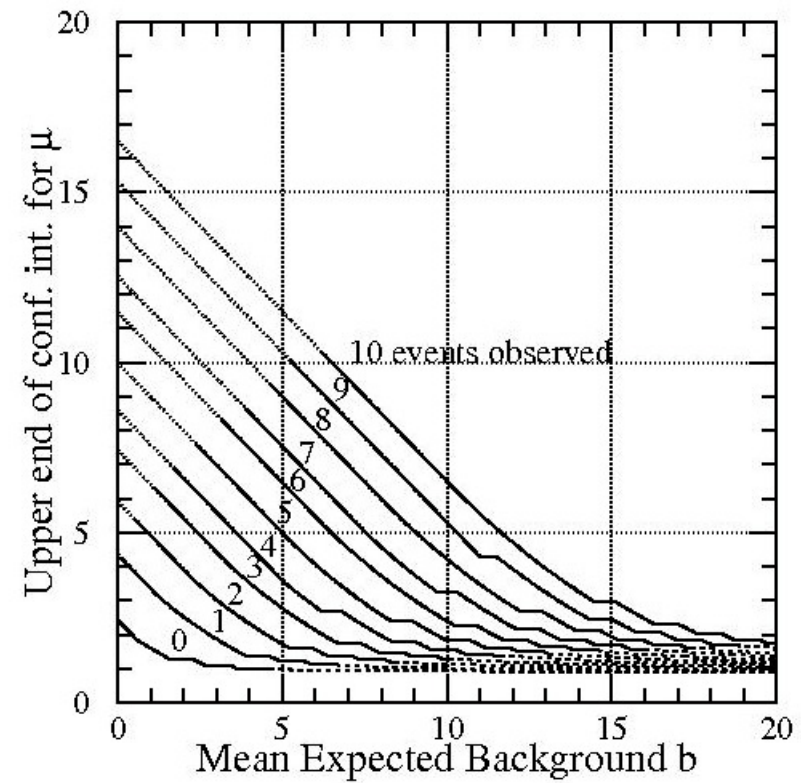
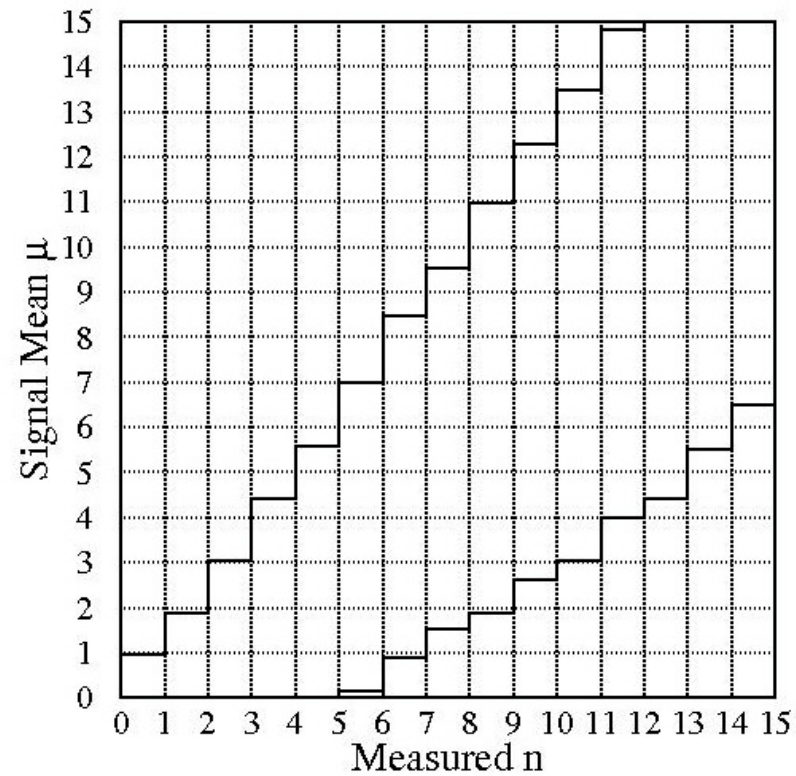
Podejście zunifikowane I

Feldman i Cousins: konstrukcja Neymana „w poziomie”

Przykład: rozkład Poissona $\mathcal{P}(n|\mu + \lambda)$ z tłem dla $\mu = 0,5$ oraz $\lambda = 3$, $1 - \alpha = 0,9$ i $\mu_{MW} = \max(0; n - \lambda)$ o włączeniu wartości n decyduje statystyka $R(n|\theta, \lambda)$

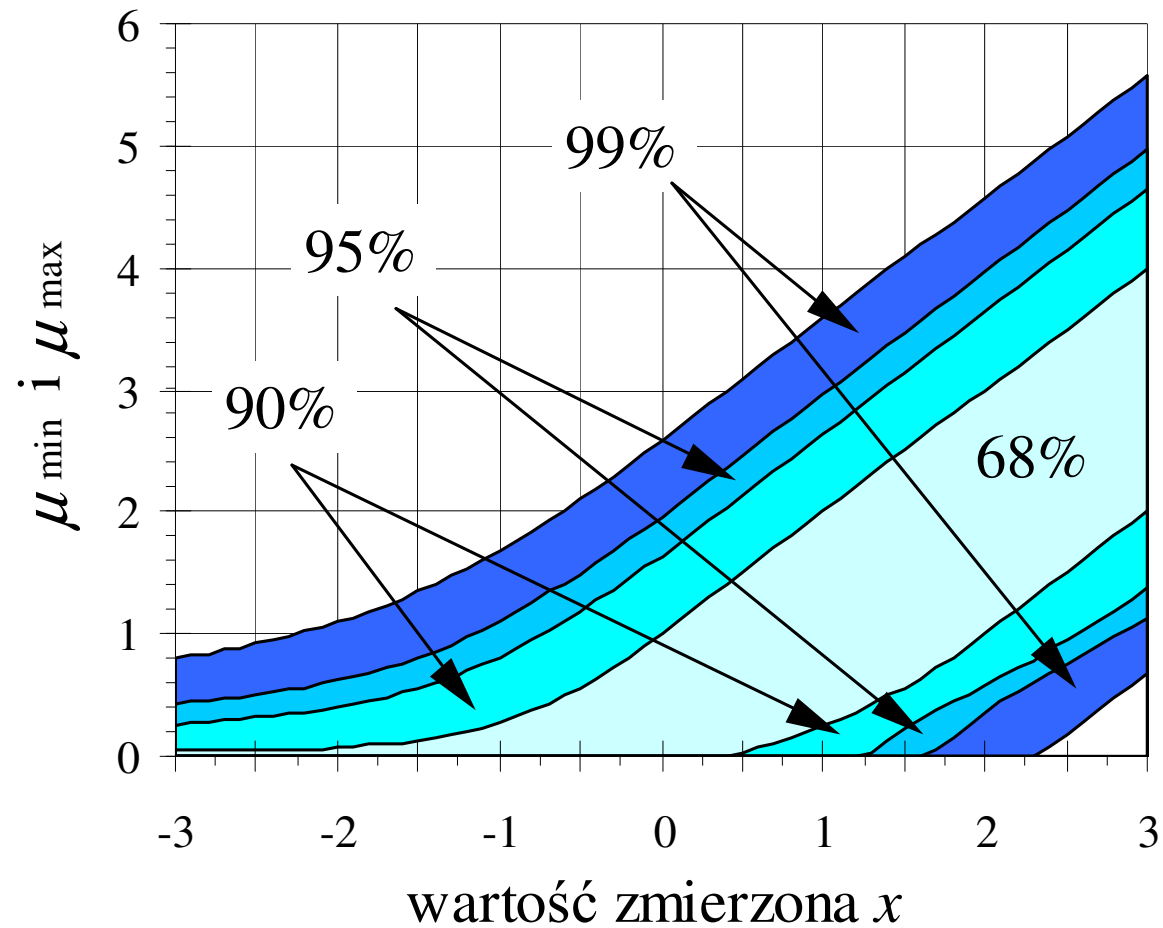
n	$P(n \mu + \lambda)$	μ_{MW}	$P(n \mu_{MW} + \lambda)$	R	ranga
0	0,030	0	0,050	0,607	6
1	0,106	0	0,149	0,708	5
2	0,185	0	0,224	0,826	3
3	0,216	0	0,224	0,963	2
4	0,189	1	0,195	0,966	1
5	0,132	2	0,175	0,753	4
6	0,077	3	0,161	0,480	7

Podejście zunifikowane II



Podjęście zunifikowane III

Granice dla wartości oczekiwanej w rozkładzie Gaussa



Podójście zunifikowane IV

Własności:

- eliminuje decyzję „przerzutnikową” (gwałcącą pokrycie)
- eliminuje нефizyczne przedziały

$N = 0, \alpha = 0,1$					
metoda	$\lambda =$	0	1	2	3
pre FC	$\mu_{\max} =$	2,3	1,3	0,3	-0,70
zunifikowana	$\mu_{\max} =$	2,44	1,61	1,26	1,08

Ale:

- jest niekonsystentne: im mniejsze tło, tym szerszy przedział
- niekiedy („tłuste” ogony) produkuje rozłączne przedziały
- nie uwzględnia niepewności tła (ocena λ z pomiaru)
- nie podaje recepty na parametry zakłócające

Prawdopodobieństwo wg. Bayesa

Definicja prawdopodobieństwa warunkowego $P(A|B)$

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$$

Twierdzenie Bayesa dla zdarzeń losowych:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

i dla zmiennych losowych:

$$f_{\theta}(\theta | \mathbf{x}) = \frac{\mathcal{L}(\mathbf{x} | \theta) z_{\theta}(\theta)}{z_x(\mathbf{x})}, \quad z_x(\mathbf{x}) = \int_{-\infty}^{\infty} \mathcal{L}(\mathbf{x} | \theta) z_{\theta}(\theta) d\theta$$

- $f_{\theta}(\theta | \mathbf{x})$ – rozkład *a posteriori*,
- $z_{\theta}(\theta), z_x(\mathbf{x})$ – rozkłady *a priori*

Przedział wiarygodności

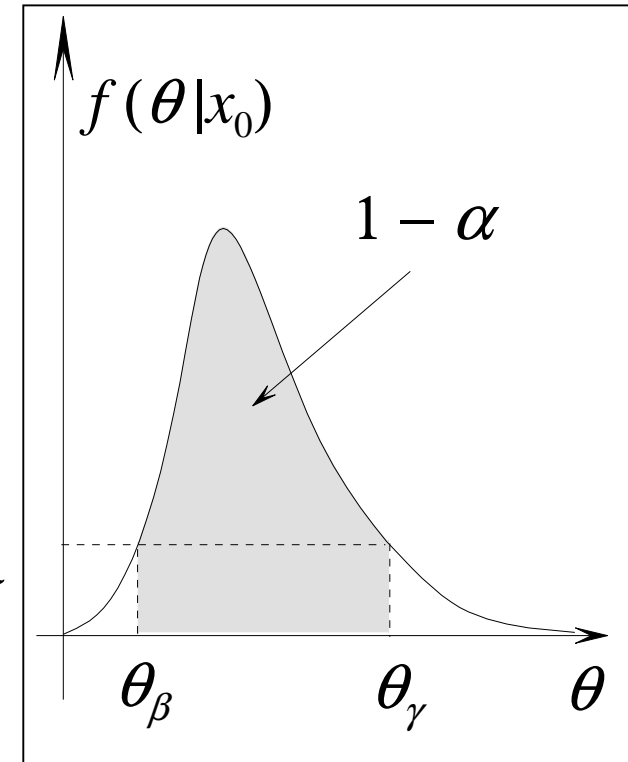
Konwencjonalny wybór – równa gęstość, ale też może być centralny, symetryczny

Efekty **systematyczne**:

$$f_{\theta}(\theta | \mathbf{x}_0) = \int_{-\infty}^{\infty} f_{\theta}(\theta | \mathbf{x}_0, \varphi) z_{\varphi}(\varphi) d\varphi$$

Pokrycie? W tym podejściu bez znaczenia

Niefizyczny przedział? Uwzględnijmy to w gęstości *a priori*



Jeden problem – **ile wynosi gęstość *a priori*?**

Gęstość *a priori*

- Laplace – zasada niedostatecznej racji, np. dla rozkładu dwumianowego $z_p(p) = \text{const}$, ale co z $z_\eta(\eta)$, jeśli $\eta = p^2$?
- Jeffreys – niezmienniczość gęstości *a priori* względem reparametryzacji modelu $\rightarrow z(\mu) = \text{const}$ oraz $z(\sigma) = 1/\sigma$ dla rozkładu Gaussa, $z(\mu) = \mu^{-1/2}$ (rozkład Poissona)
- Jaynes – zasada maksymalnej entropii \rightarrow problem wariacyjny

$$S(\theta) = \int z_\theta(\theta) \ln z_\theta(\theta) d\theta = \max(z_\theta(\theta))$$

- Bernardo – gęstość referencyjna \rightarrow przyrost informacji (odległość Kullbacka-Leiblera) \rightarrow problem wariacyjny

$$\int f_\theta(\theta|\mathbf{x}) \ln \frac{f_\theta(\theta|\mathbf{x})}{z_\theta(\theta)} d\theta$$

- Typowy fizyk – to, na co pozwala *mu* jego wiedza i jego doświadczenie życiowe (czyli najczęściej: $z_\theta(\theta) = \text{const}$)

Metoda Bayesa przy pracy

Rozkład Poissona: $\mathcal{P}(n|\mu + \lambda) = \mathcal{L}(n|\mu, \lambda)$ z tłem, $z_\mu(\mu) = \text{const}$

$$f_\mu(\mu|n, \lambda) = \frac{\mathcal{P}(n|\mu + \lambda)}{\sum_{k=0}^N \mathcal{P}(k|\lambda)}$$

$$1 - \alpha = \int_0^\mu f_\mu(\mu'|n, \lambda) d\mu' \Rightarrow \alpha = \sum_{n=0}^N f_\mu(\mu|n, \lambda)$$

$N = 0, \alpha = 0,1$					
metoda	$\lambda =$	0	1	2	3
pre FC	$\mu_{\max} =$	2,30	1,30	0,30	-0,70
zunifikowana	$\mu_{\max} =$	2,44	1,61	1,26	1,08
bayesowska	$\mu_{\max} =$	2,30	2,30	2,30	2,30

Która metoda?

- Podejście klasyczne stosuje żelazną logikę i metody nie do zakwestionowania w rozwiązaniu problemu, w którym nikt nie jest zainteresowany
- Podejście bayesowskie używa „szemranych” metod, aby uzyskać odpowiedź na pytanie, w którym wszyscy są zainteresowani

Która metoda?

Różnice natury psychologiczne – dwa różne światy

Zalecenie generalne: konsument/producent musi być świadom swego wyboru, a więc:

- analiza bayesowska czy klasyczna
- jeśli bayesowska – postać rozkładów *a priori*
- który z wariantów, jak „zmodernizowany”
- klarowna procedura ze wszystkimi szczegółami
- apel o postać funkcji wiarygodności
- poparcie dla dwóch klas analiz

Ograniczenia na masę cząstki Higgsa I

Funkcja wiarygodności:

N_c – liczba procesów fizycznych

$N_{i,bin}$ – liczba przedziałów histogramu w i -tym procesie

$$\mathcal{L}(\mathbf{n} | R, \boldsymbol{\theta}) = \prod_{i=1}^{N_c} \prod_{j=1}^{N_{i,bin}} \frac{\eta_{ij}^{n_{ij}}}{n_{ij}!} e^{-\eta_{ij}}, \quad \eta_{ij} = R\mu_{ij}(\boldsymbol{\theta}) + \lambda_{ij}(\boldsymbol{\theta})$$

R – sensitivity factor = σ/σ_{SM}

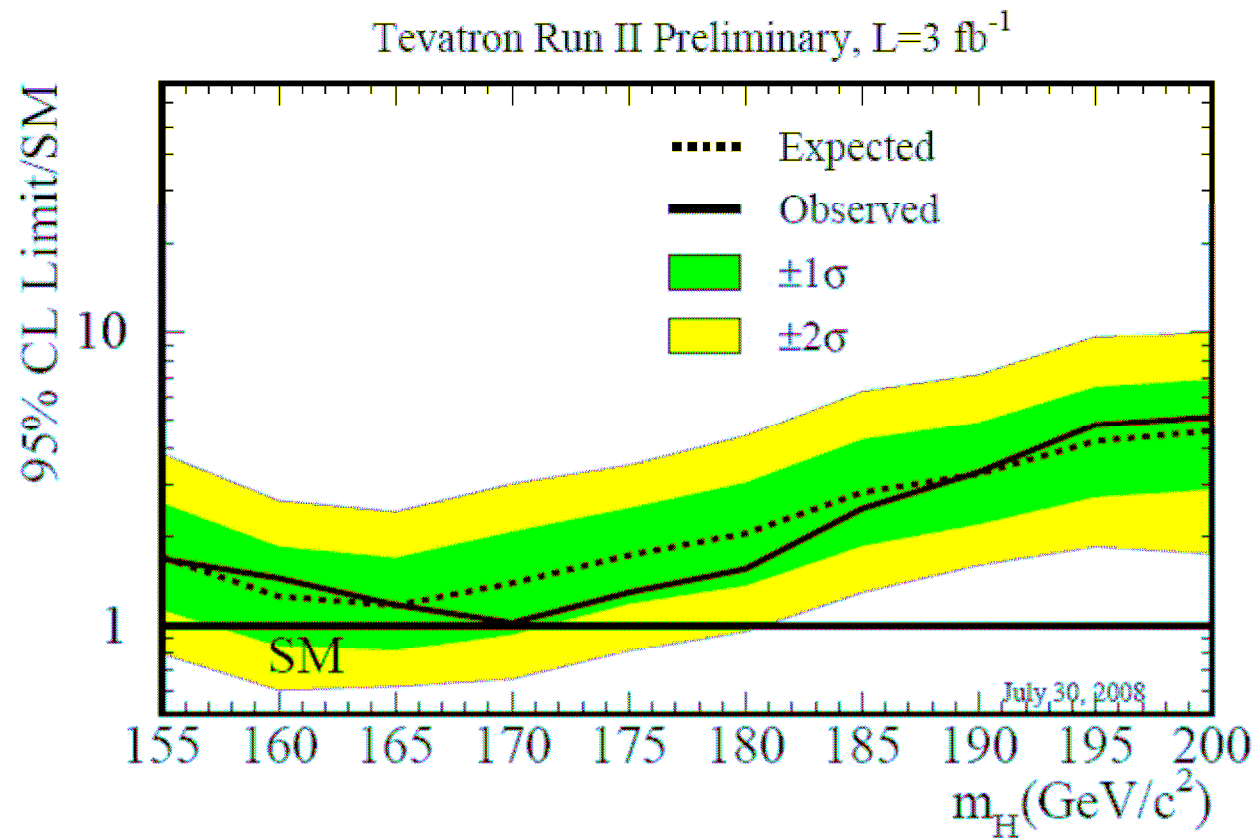
Rozkłady *a priori*:

$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \boldsymbol{\theta}_0, \boldsymbol{\Sigma})$ – gaussowski kształt z korelacjami

$z_R(R)$ – rozkład jednostajny

$$f_R(R | \mathbf{n}) = \frac{1}{z_n(\mathbf{n})} \int \mathcal{L}(\mathbf{n} | R, \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \boldsymbol{\theta}_0, \boldsymbol{\Sigma}) z_R(R) d\boldsymbol{\theta}$$

Ograniczenia na masę cząstki Higgsa II



Podsumowanie

„If your experiment needs statistics,
you ought to have done a **better** experiment.”

Ernest Rutherford