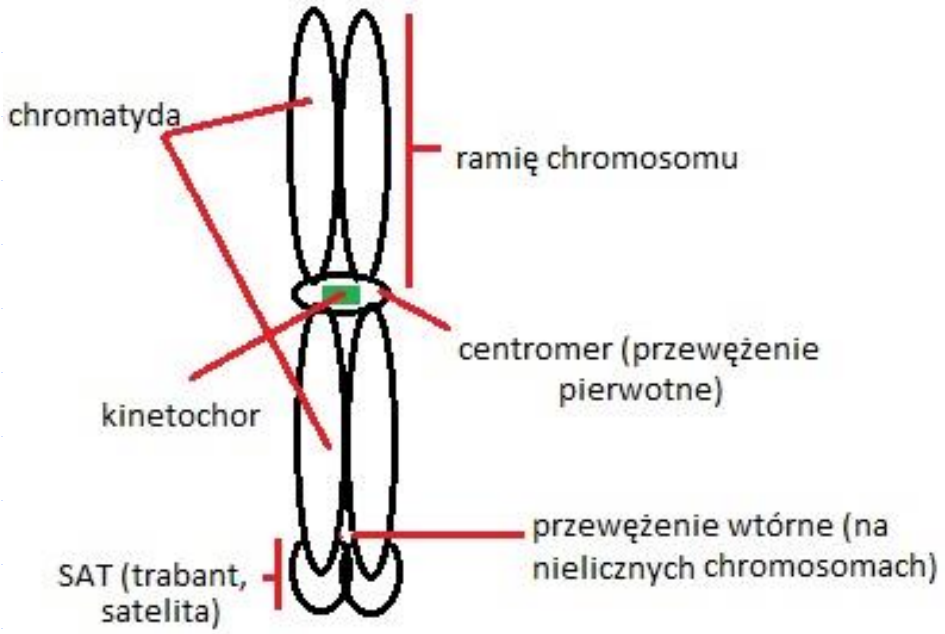
BIOFIZYKA RADIACYJNA
wykład #7

Krzysztof Wojciech Fornalski

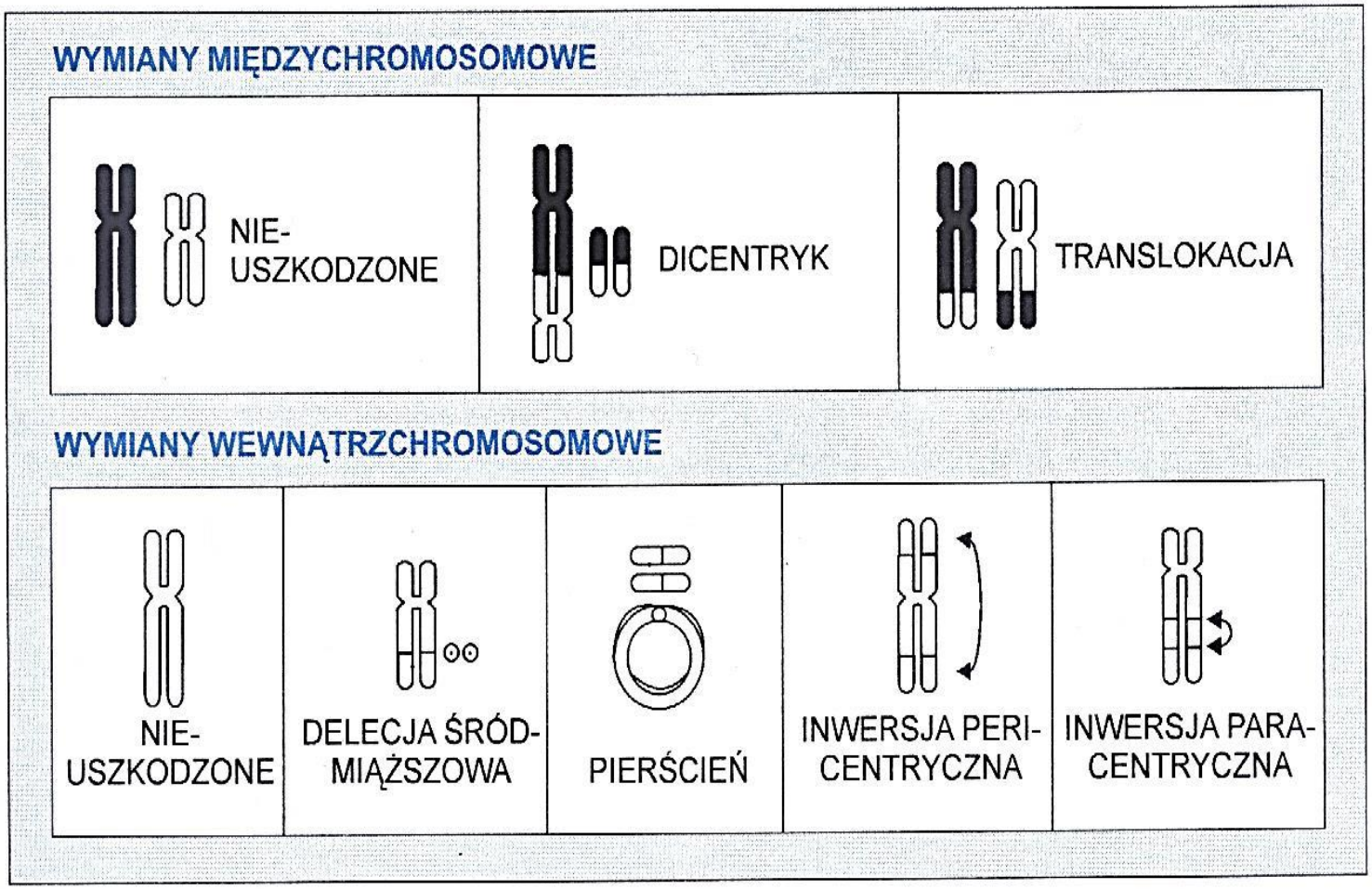
Wydział Fizyki, Politechnika Warszawska

PRZYPOMNIENIE Z POPRZEDNIEGO WYKŁADU

Budowa chromosomu



Aberracje chromosomowe



Rys. 2. *Schemat wymian wewnątrz- i międzychromosomowych obserwowanych w komórkach narażonych na działanie promieniowania jonizującego*

BIODOZYMETRIA (DOZYMETRIA BIOLOGICZNA)

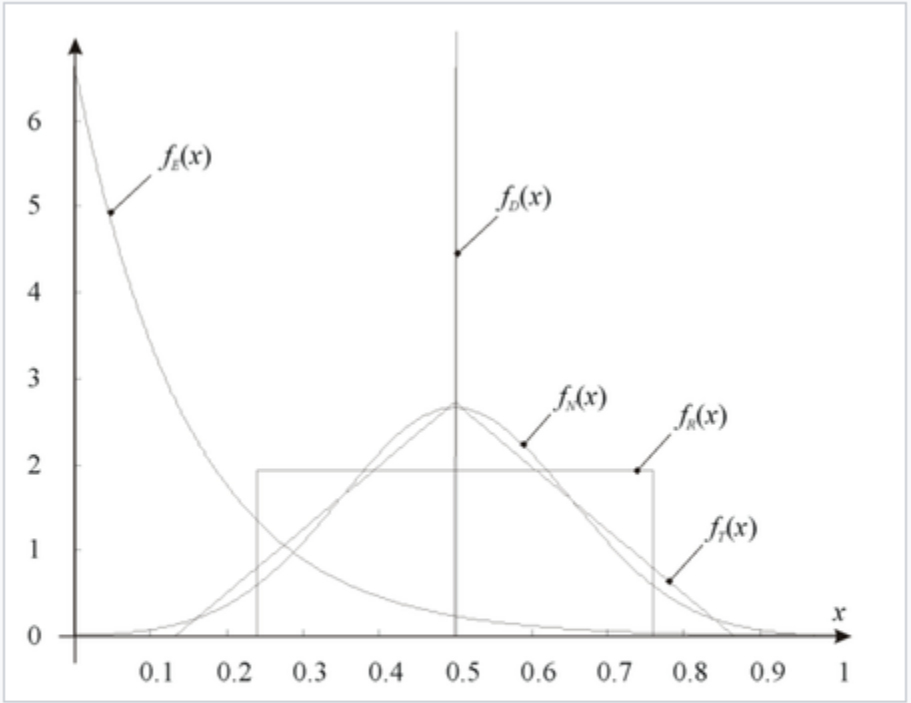
- BIODOZYMETRIA CYTOGENETYCZNA DICENTRYKÓW
- ANALIZA BAYESOWSKA

DYGRESJA: ZAAWANSOWANE METODY STATYSTYCZNE

- ROZKŁAD PRAWDOPODOBIEŃSTWA
- FUNKCJA WIARYGODNOŚCI
- BAYESOWSKA ANALIZA DANYCH

Rozkład (funkcja) prawdopodobieństwa, PDF

- Nie definiujemy prawdopodobieństwa klasycznie (częstościowo), lecz poprzez funkcję prawdopodobieństwa
- Termin „funkcja prawdopodobieństwa” jest formalnie zarejestrowany do rozkładów dyskretnych
- Dla rozkładów ciągłych – formalnie „funkcja gęstości rozkładu prawdopodobieństwa” (tzw. PDF – probability density function)



Wybrane rozkłady gęstości prawdopodobieństwa:
$f_N(x)$ – rozkład normalny,
$f_E(x)$ – rozkład wykładniczy,
$f_R(x)$ – rozkład jednostajny,
$f_T(x)$ – rozkład trójkątny,
$f_D(x)$– rozkład delty Diraca dla zmiennej pewnej.

Funkcja wiarygodności (likelihood, L)

wprowadzimy najpierw tzw. **funkcję wiarygodności**. Niech będzie dana próbka losowa x_i o liczebności n z rozkładu $f(x;\theta)$, gdzie x jest zmienną losową określoną, dla ustalenia uwagi, na całej osi rzeczywiste, a θ parametrem określającym rozkład. Funkcją $\mathscr{L}$ wiarygodności dla próbki x_i nazywamy wielkość:

$$\mathscr{L}(\vec{\mathbf{x}};\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Podobnie, dla losowej próbki k_i ($i = 1, 2, \ldots, n$) zmiennych dyskretnych z rozkładu $P_k(\theta)$, zakładając dla ustalenia uwagi, że $k = 0, 1, 2, \ldots$, funkcję wiarygodności definiujemy jako:

$$\mathscr{L}(\vec{\mathbf{k}};\theta) = \prod_{i=1}^{n} P_{k_i}(\theta).$$

Należy zwrócić uwagę na to, że formalnie funkcja wiarygodności wygląda jak łączna funkcja gęstości rozkładu. I taką łączną funkcją gęstości jest ona tak długo, jak długo wielkości x_i oraz k_i w obu wyrażeniach są *zmiennymi* losowymi. Wielokrotnie, w dalszej części wykładu, spotkamy się z sytuacjami, gdy wielkości x_i oraz k_i to faktycznie wyniki pomiaru, a więc ściśle określone liczby, a nie zmienne. Wtedy wielkość $\mathscr{L}$ nie jest funkcją gęstości zmiennych losowych – jest to zwykła, matematyczna funkcja, zależna tylko i wyłącznie od parametru θ (jednego lub wielu). W literaturze statystycznej utarła się i bardzo głęboko zakorzeniła się tradycja wymiennego stosowania terminu *funkcja wiarygodności* dla obu tych sytuacji.

W praktyce częściej korzystamy z logarytmu z L, bo zarówno L jak i ln L osiągają ekstremum dla tych samych wartości, a obliczenia są prostsze (logarytm z iloczynu to suma logarytmów)

Funkcja wiarygodności (likelihood, L)

wprowadziliśmy funkcję wiarygodności:

$$\mathscr{L}(\vec{\mathbf{x}};\theta)=\prod_{i=1}^{n} f(x_i;\theta)$$

jako łączną funkcję gęstości, w której w miejsce zmiennych losowych x_i podstawiamy wartości uzyskane w wyniku pomiaru, a więc w wyniku pobierania próbki prostej z rozkładu $f(x;\theta)$. Zakładamy, że postać funkcji gęstości $f(x;\theta)$ jest znana, nie znamy jednak wartości parametru θ. W takie sytuacji, funkcja wiarygodności staje się funkcją tego i tylko tego parametru. Zgodnie z **zasadą (metodą) największej wiarygodności**, za wartość nieznanego parametru θ winniśmy wybrać taką liczbę $\hat{\theta}$, dla której funkcja wiarygodności osiąga maksimum:

$$\boxed{\mathscr{L}(\vec{\mathbf{x}};\hat{\theta})=\max}.$$

Żądanie maksymalnej wartości funkcji wiarygodności sprowadza się do znalezienia pierwiastka $\hat{\theta}$ równania:

$$\frac{\partial}{\partial\theta}\mathscr{L}(\vec{\mathbf{x}};\theta)=0 \quad \text{przy warunku:} \quad \frac{\partial^2}{\partial\theta^2}\mathscr{L}(\vec{\mathbf{x}};\theta)\Big|_{\theta=\hat{\theta}}<0\,.$$

Zazwyczaj funkcja największej wiarygodności ma jedno maksimum. Gdy nie jest to prawda, musimy poszukać dodatkowych informacji, pozwalających dokonać jednoznacznego wyboru.

Ponieważ funkcja $\mathscr{L}$ jak i jej logarytm osiągają maksimum dla tej samej wartości argumentu, estymator największej wiarygodności można często znaleźć z bardziej praktycznego równania:

$$\frac{\partial}{\partial\theta}\ln\mathscr{L}(\vec{\mathbf{x}};\theta)=\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)=0\,, \quad \text{przy warunku:} \quad \sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(x_i;\theta)\Big|_{\theta=\hat{\theta}}<0\,.$$

Dla ogólnego przypadku z k nieznanymi parametrami θ_i, musimy rozwiązać układ k równań:

$$\frac{\partial}{\partial\theta_j}\ln\mathscr{L}(\vec{\mathbf{x}};\vec{\boldsymbol{\theta}})=\sum_{i=1}^{n}\frac{\partial}{\partial\theta_j}\ln f(x_i;\vec{\boldsymbol{\theta}})=0\,, \quad \text{dla:} \quad j=1,2,\ldots,k$$

Szacowanie niepewności parametru opisywanego funkcją wiarygodności

W języku funkcji L, a więc łącznej funkcji gęstości, ograniczenie od dołu na wartość wariancji nieobciążonego estymatora zadana jest przez nierówność określoną **twierdzeniem Rao–Cramera**:

$$V\left(\hat{\theta}\right) \geq V_{\min}\left(\hat{\theta}\right) = \frac{1}{\int L\left(\vec{\mathbf{x}};\theta\right)\left(\frac{\partial}{\partial\theta}\ln L\left(\vec{\mathbf{x}};\theta\right)\right)^2 d\vec{\mathbf{x}}},$$

i podobnie dla zmiennej dyskretnej:

$$V\left(\hat{\theta}\right) \geq V_{\min}\left(\hat{\theta}\right) = \frac{1}{\sum_{\vec{\mathbf{k}}} L\left(\vec{\mathbf{k}};\theta\right)\left(\frac{\partial}{\partial\theta}\ln L\left(\vec{\mathbf{k}};\theta\right)\right)^2},$$

Możemy przyjąć, że wariancja V to kwadrat niepewności σ^2

Twierdzenie Bayesa

Podstawy rozumowania bayesowskiego sięgają podstaw teorii prawdopodobieństwa, a mianowicie prawdopodobieństwa warunkowego zdarzenia *X* jeśli zachodzi zdarzenie *Y*. Prawdopodobieństwo to dane jest znanym wzorem: $P(X/Y) = P(X \cap Y) / P(Y)$. Z racji tego, iż prawdopodobieństwa iloczynów zdarzeń są sobie równe, $P(X \cap Y) = P(Y \cap X)$, zapisać można:

$$P(X \mid Y, I) = \frac{P(Y \mid X, I) \times P(X \mid I)}{P(Y \mid I)}$$

gdzie *I* oznacza posiadaną wcześniej informację o obiekcie (jeśli ona istnieje).

Powyższy wzór stanowi treść twierdzenia Thomasa Bayesa (1702-1761) i wiąże prawdopodobieństwa warunkowe $P(X/Y,I)$ oraz $P(Y/X,I)$ przy posiadaniu dodatkowej informacji *I*. W ogólności parametr *I* określa wszelkie dane, które są niezależne od *X* oraz *Y*, a które mówią o przedmiocie badań, np. wyniku poprzedniego eksperymentu.

Twierdzenie Bayesa

Jeśli przez *X* i *Y* rozumiemy zmienne fizyczne, które możemy wielokrotnie zmierzyć, sens prawdopodobieństwa może zostać sprowadzony do prawdopodobieństwa otrzymania takich właśnie wartości w nieskończonej liczbie eksperymentów powtarzanych w identycznych warunkach. Jednakże, jeśli któraś z tych wielkości miałaby oznaczać np. hipotezę, takie podejście, zwane „częstotliwościowym", traci sens. Wówczas dobrym przykładem zastosowania twierdzenia Bayesa jest użycie równania w tzw. interpretacji „subiektywnej". Przyjmując za *X* pewną teorię (hipotezę) fizyczną *T,* a za *Y* dane eksperymentalne *E*, określić można stan naszej wiedzy na temat prawdziwości danej teorii w świetle posiadanych wyników doświadczalnych i wcześniejszej wiedzy. Wówczas czynnik *P(E|T,I)* oznacza funkcję wiarygodności (L) mówiącą o tym, jak dobrze dane eksperymentalne *E* odpowiadają teorii *T*, natomiast człon *P(T|I)* określa stopień zaufania dla danej teorii, w skali ciągłej od 0 do 1, zanim przeprowadzono eksperyment. Tę wiedzę aprioryczną określać będziemy słowem *prior* jak w oryginalnej literaturze angielskojęzycznej. Człon z mianownika, *P(E|I)*, jest w praktyce członem normalizacyjnym.

Twierdzenie Bayesa

Ostatecznie otrzymujemy:

$$P(T \mid E, I) = \frac{P(E \mid T, I) \times P(T \mid I)}{P(E \mid I)} \propto P(E \mid T, I) \times P(T \mid I)$$

Funkcja wiarygodności może przyjąć postać konkretnego rozkładu (np. Gaussa), a *prior* być funkcją z góry nakładającą pewne obostrzenia np. na parametry teorii, która powinna opisać dany eksperyment. Równanie powyższe pokazuje ponadto, w jaki sposób zmienia się zaufanie do teorii *T* w wyniku przeprowadzenia eksperymentu i otrzymania wyników *E*.

Rozumowanie bayesowskie

- Na podstawie twierdzenia Bayesa zapisać więc można ogólne równanie na prawdopodobieństwo wynikowe (posterior)

$$\text{Posterior} \sim L \cdot \text{prior}$$

- gdzie L to likelihood (funkcja wiarygodności), zaś prior to funkcja aprioryczna
- Zarówno L jak i prior są rozkładami prawdopodobieństwa (PDF)

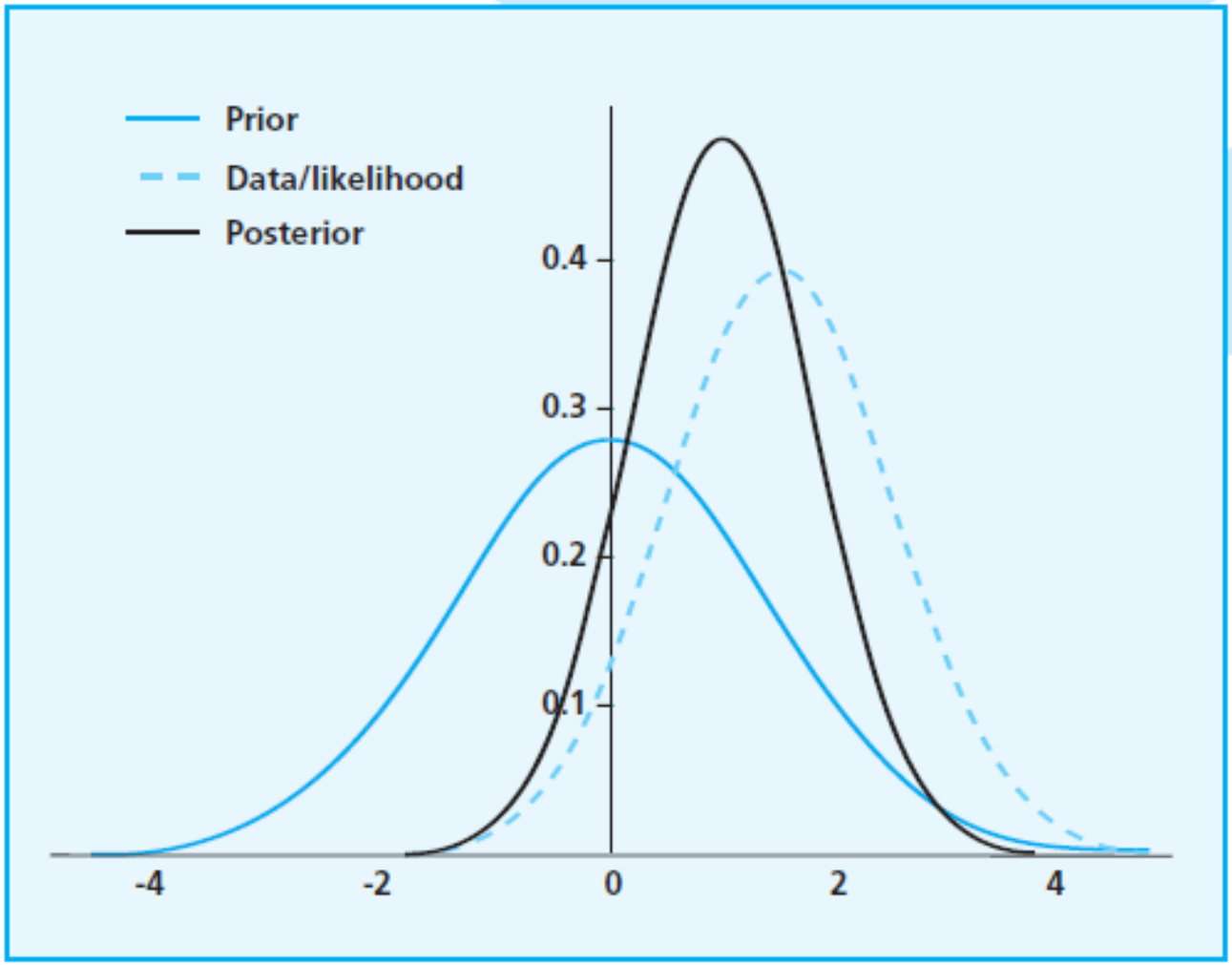
Rozumowanie bayesowskie

$$\text{Posterior} \sim \text{Likelihood} \times \text{Prior}$$

Powyższe równanie należy interpretować (i używać!) następująco: szukamy rozkładu p-twa jakiegoś parametru, np. wyniku w rzucie kostką. Rozkład ten jest naszym wynikiem (funkcja posterior). Wynik ten otrzymujemy wymnażając dwie inne funkcje: 1) wiarygodności (likelihood), która mówi o obiektywnych kryteriach (danych) wynikających np. z eksperymentu (np. p-two wyrzucenia każdego z 6 wyników na kostce jest równe), 2) prior, która określa dodatkową wiedzę (lub wyniki innych eksperymentów), np. że naroża kostki zostały spiłowane.

$$POSTERIOR\ PROB. = LIKELIHOOD\ PROB. \times PRIOR\ PROB.$$

$$P(\Theta) = L(\Theta) \cdot p(\Theta)$$



Mamy posterior PDF – i co dalej?

- Nasz wynik – posterior (P) – czyli funkcja rozkładu prawdopodobieństwa, posiada zazwyczaj jedno maksimum, które odpowiada najbardziej prawdopodobnej wartości szukanego parametru (Θ)
- Szukamy tego parametru szukając maksimum rozkładu P:

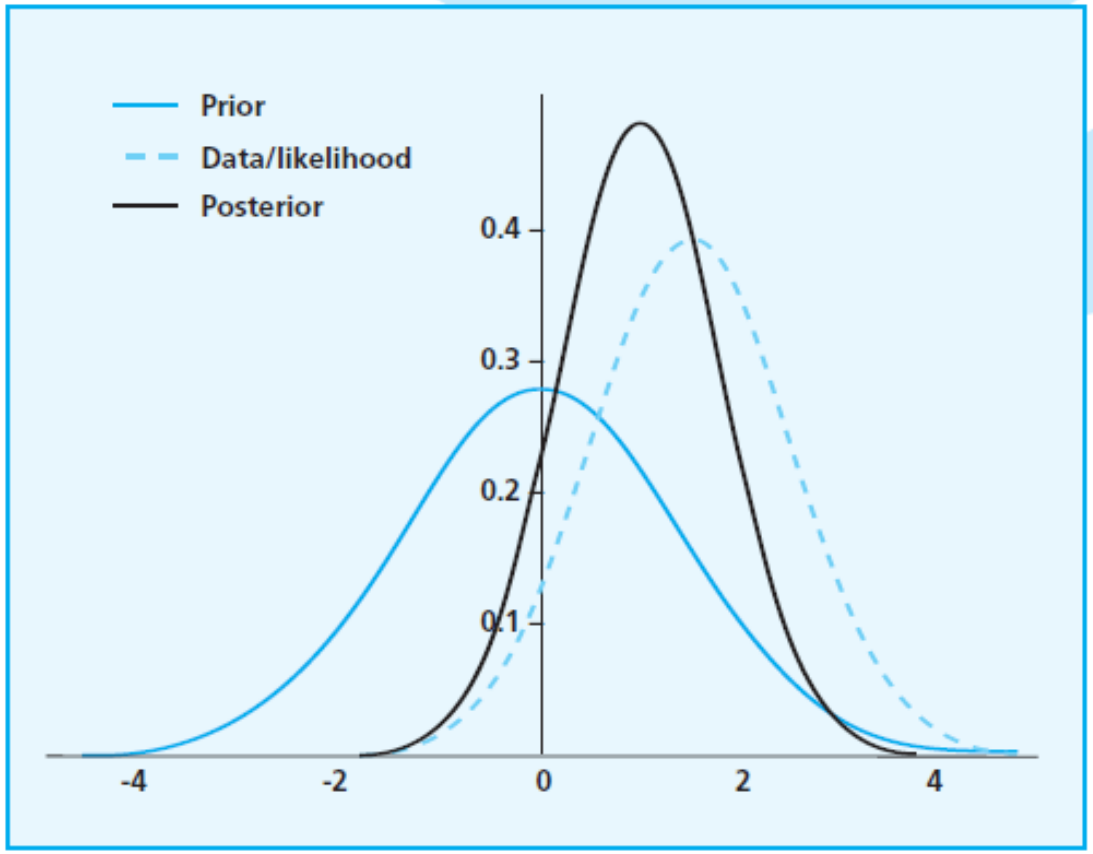
$$\frac{dP(\theta)}{d\theta} = \frac{d\,[\,L(\theta)\cdot p(\theta)\,]}{d\theta} = 0$$

Różnica między metodą bayesowską a metodą największej wiarygodności

- Kilka slajdów wcześniej opisywaliśmy analogiczne działanie dla samej funkcji wiarygodności (szukaliśmy jej maksimum aby znaleźć najbardziej prawdopodobny parametr)
- W metodzie bayesowskiej mamy dodatkową funkcję prior, którą mnożymy z f-cją wiarygodności
- Pojęcie funkcji prior, jej prawidłowe sformułowanie, jest tu krytycznym punktem

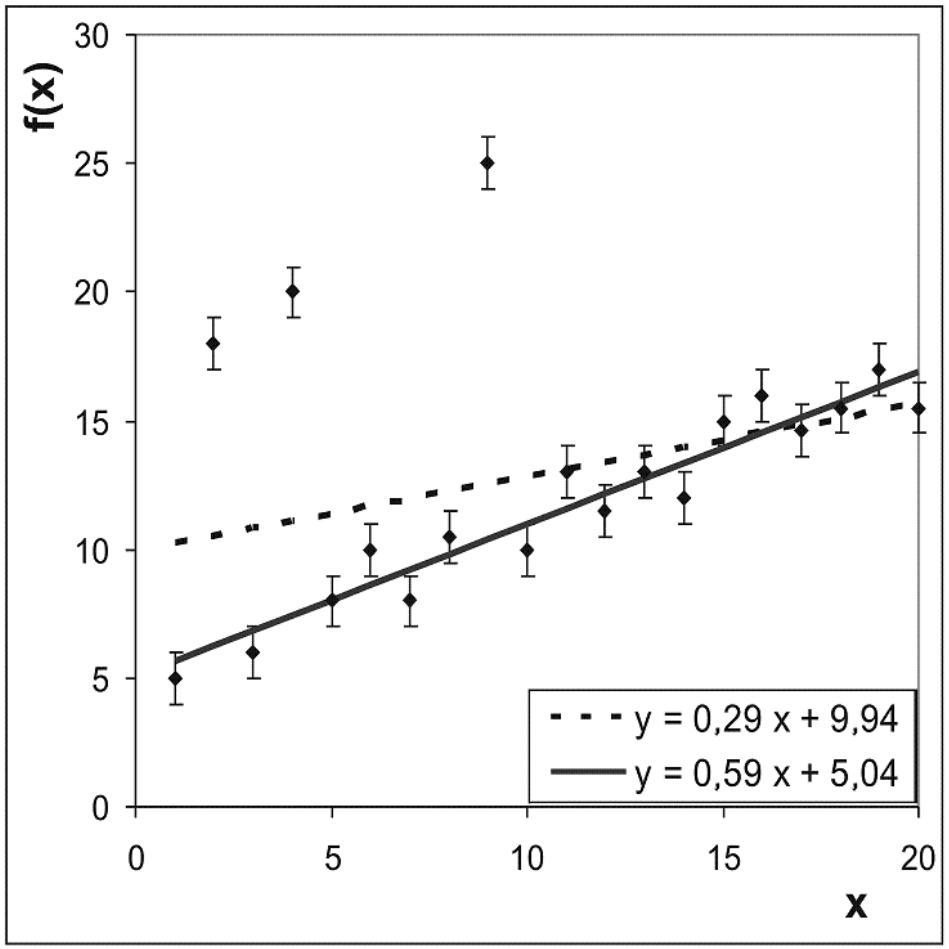
Przykład: uczenie się

- Wynik kolejnego eksperymentu/danej jest funkcją L
- Zaś wyniki wszystkich poprzednich: priorem
- I tak posterior i-tego działania staje się priorem dla i+1 działania



Przykład – regresja liniowa: porównanie metody najmniejszych kwadratów a dopasowania bayesowskiego (odpornego na punkty wybite)

- W dopasowaniu bayesowskim funkcja wiarygodności to rozkład Gaussa (jak w metodzie najmniejszych kwadratów), zaś prior kwestionuje poprawność istniejących niepewności i rozciąga je powodując, że ich wkład do finalnego posterioru (iloczynu) jest nieznaczący

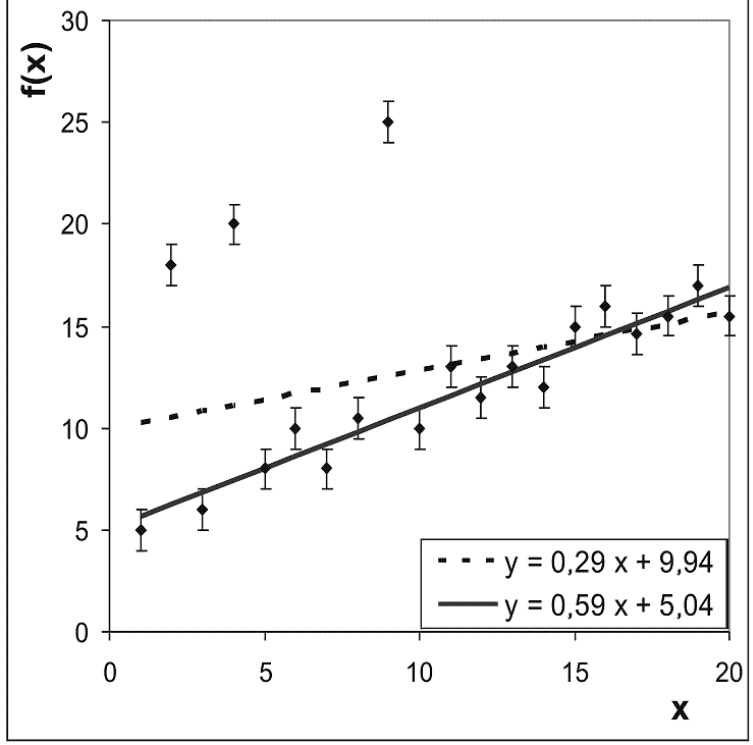


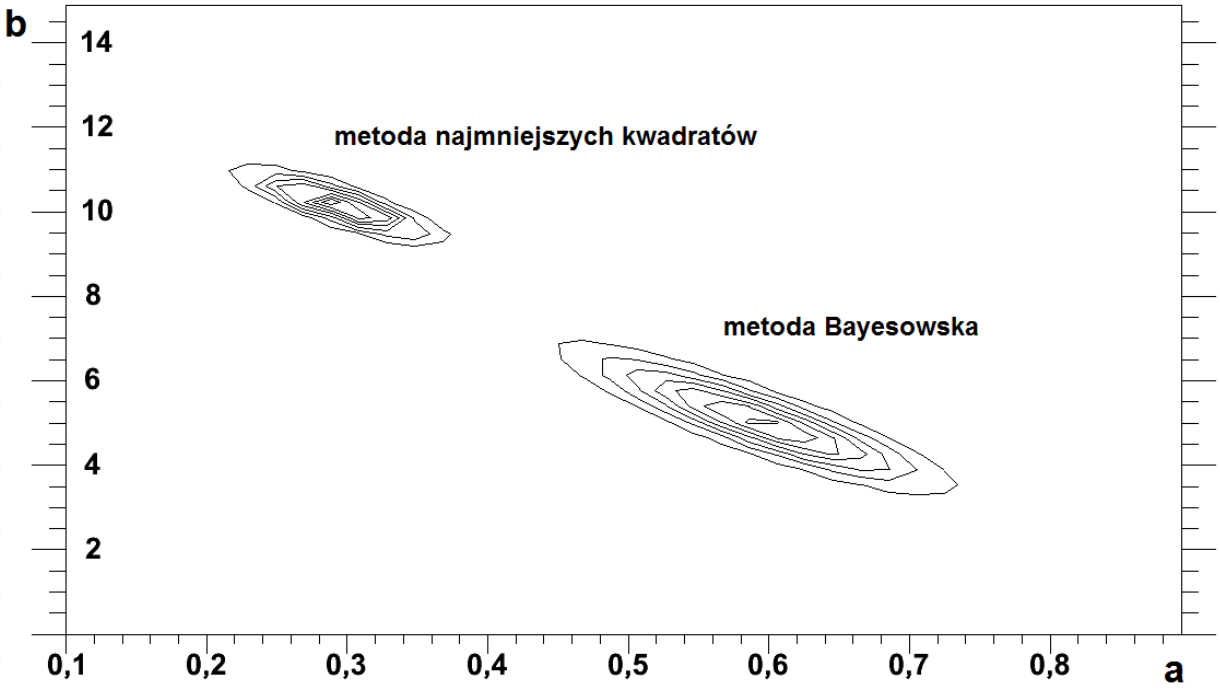
$$P = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-(T-E)^2/(2\sigma^2)\right\}p(\sigma|I)$$

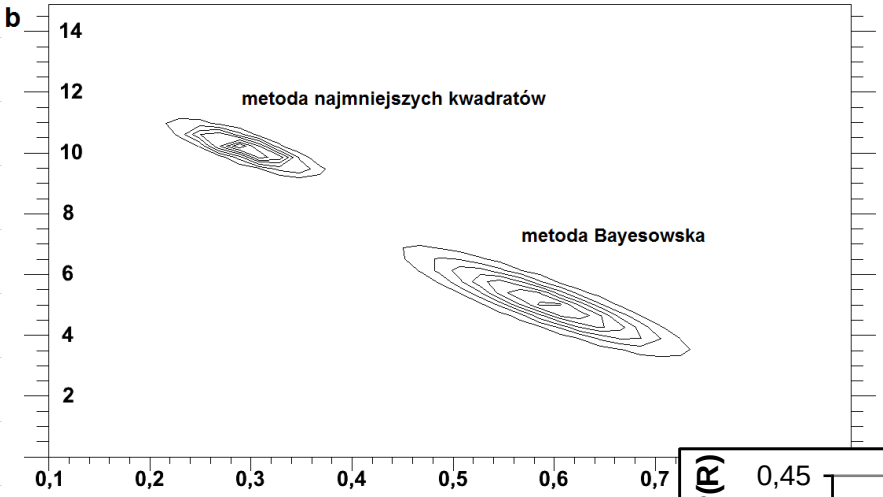
gdzie prior: $p(\sigma) = \frac{\sigma_0}{\sigma^2}$

Regresja liniowa – c.d.

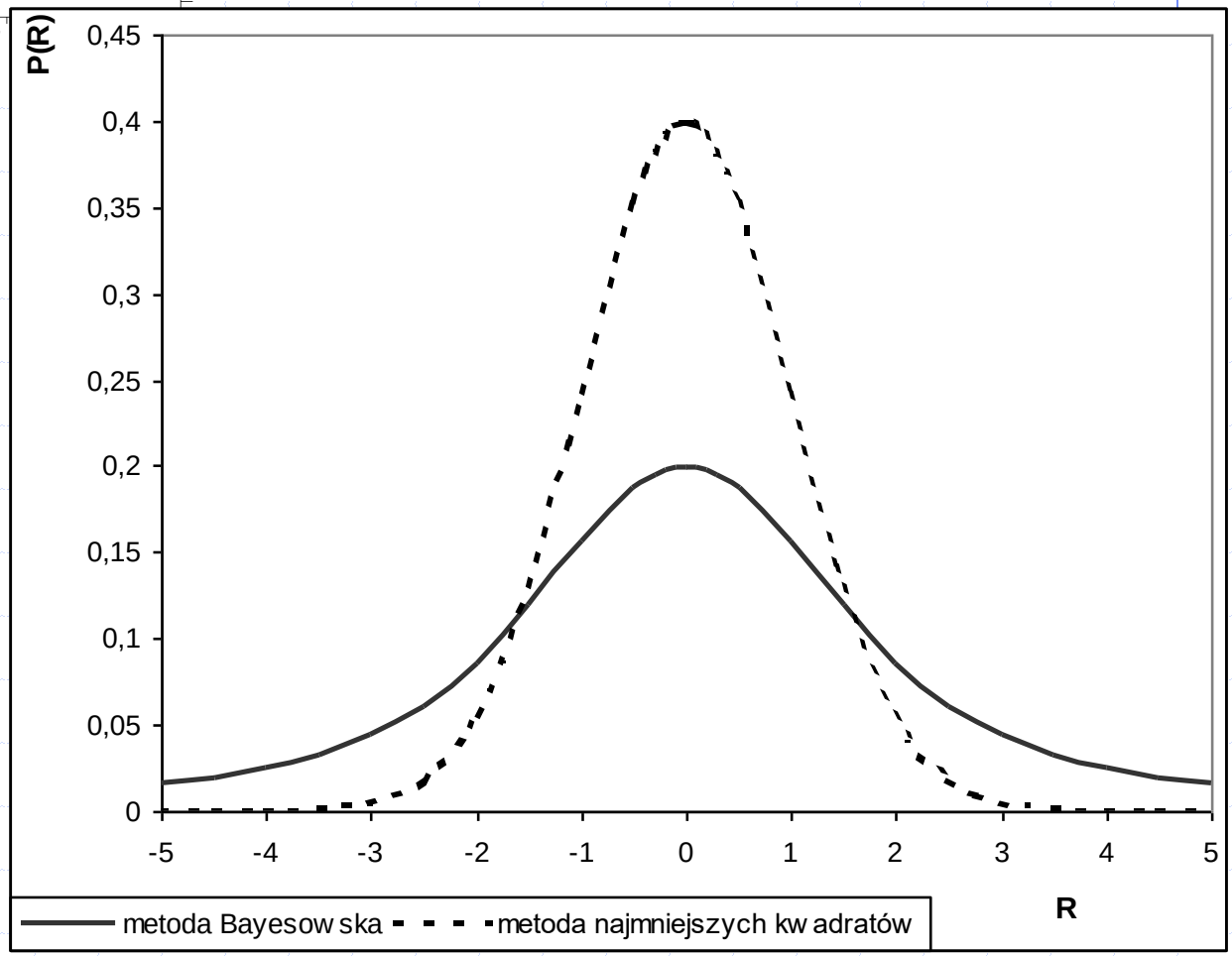
- Przedstawiona metoda jest przykładem tzw. robust regression analysis, czyli metody nieczułej na punkty wybite
- Oczywiście skutkuje to większymi niepewnościami otrzymanych wyników (parametrów dopasowania prostej)



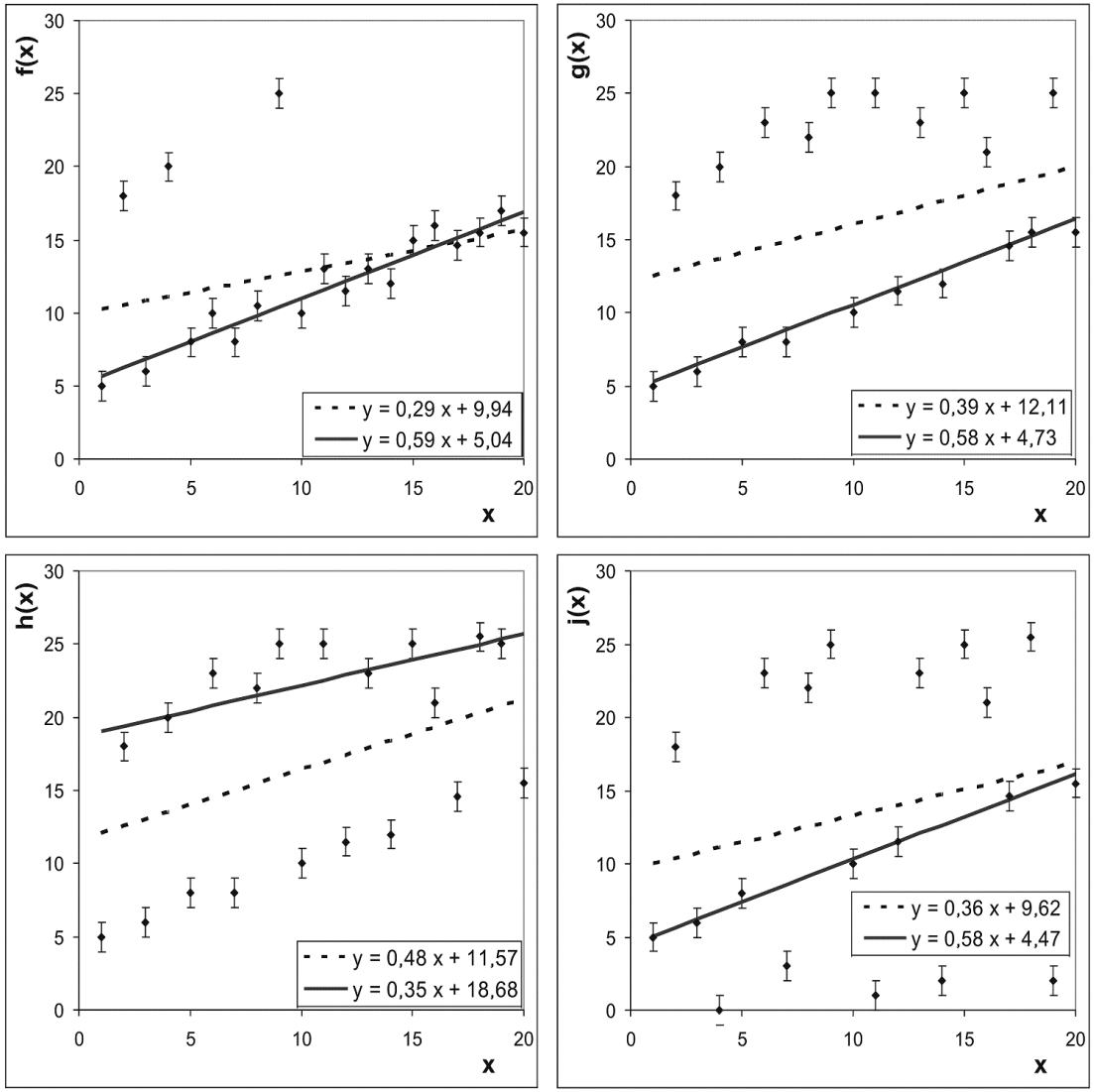




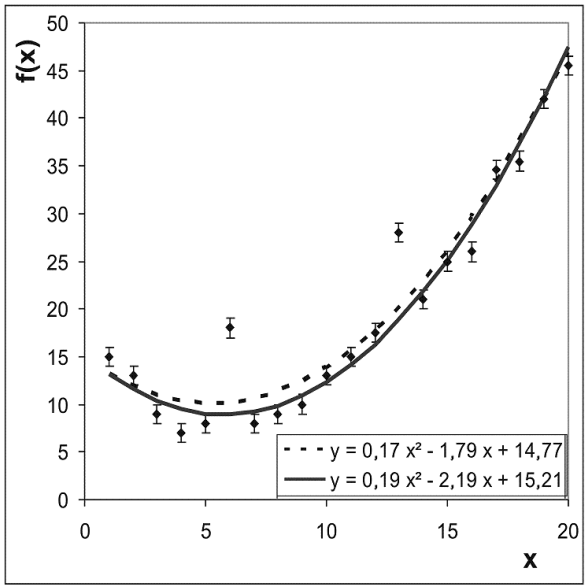
- Rozkład p-twa dla jednego punktu – porównanie metody bayesowskiej i najmniejszych kwadratów; prior rozmywa nam ogony

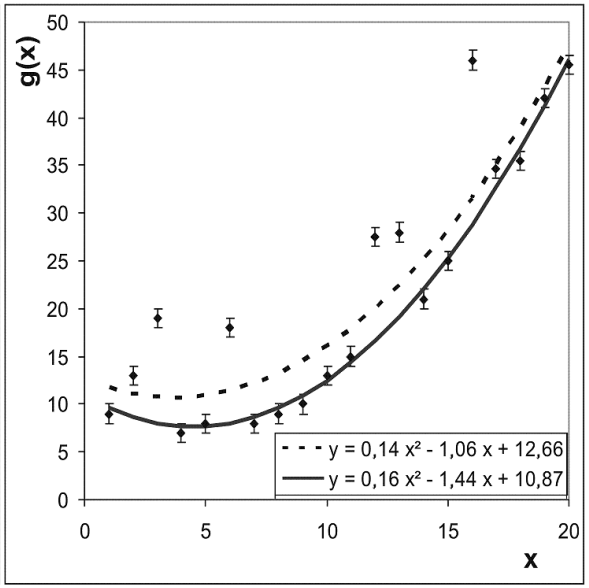


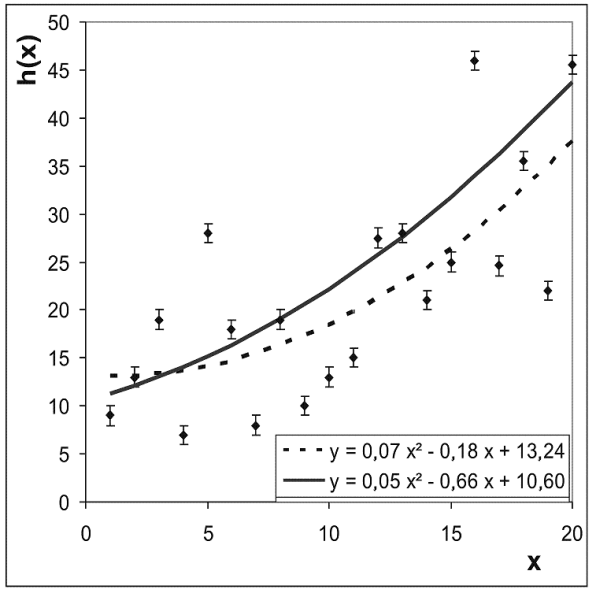
Dajmy jeszcze więcej punktów wybitych (tzw. outlier'ów)

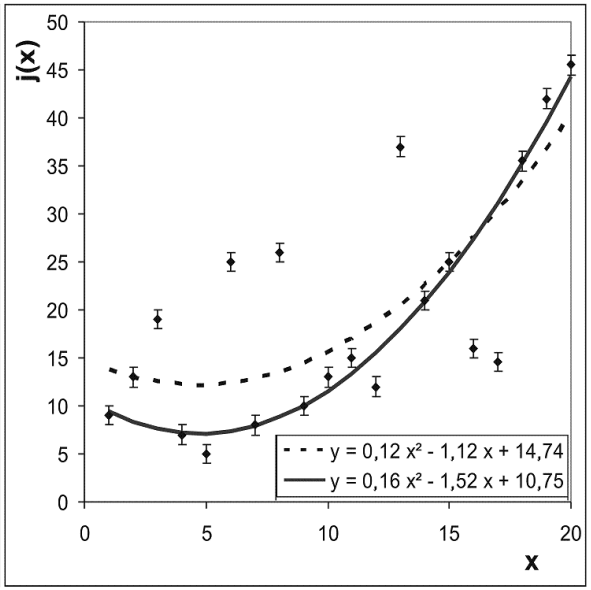


Regresja dla wielomianu

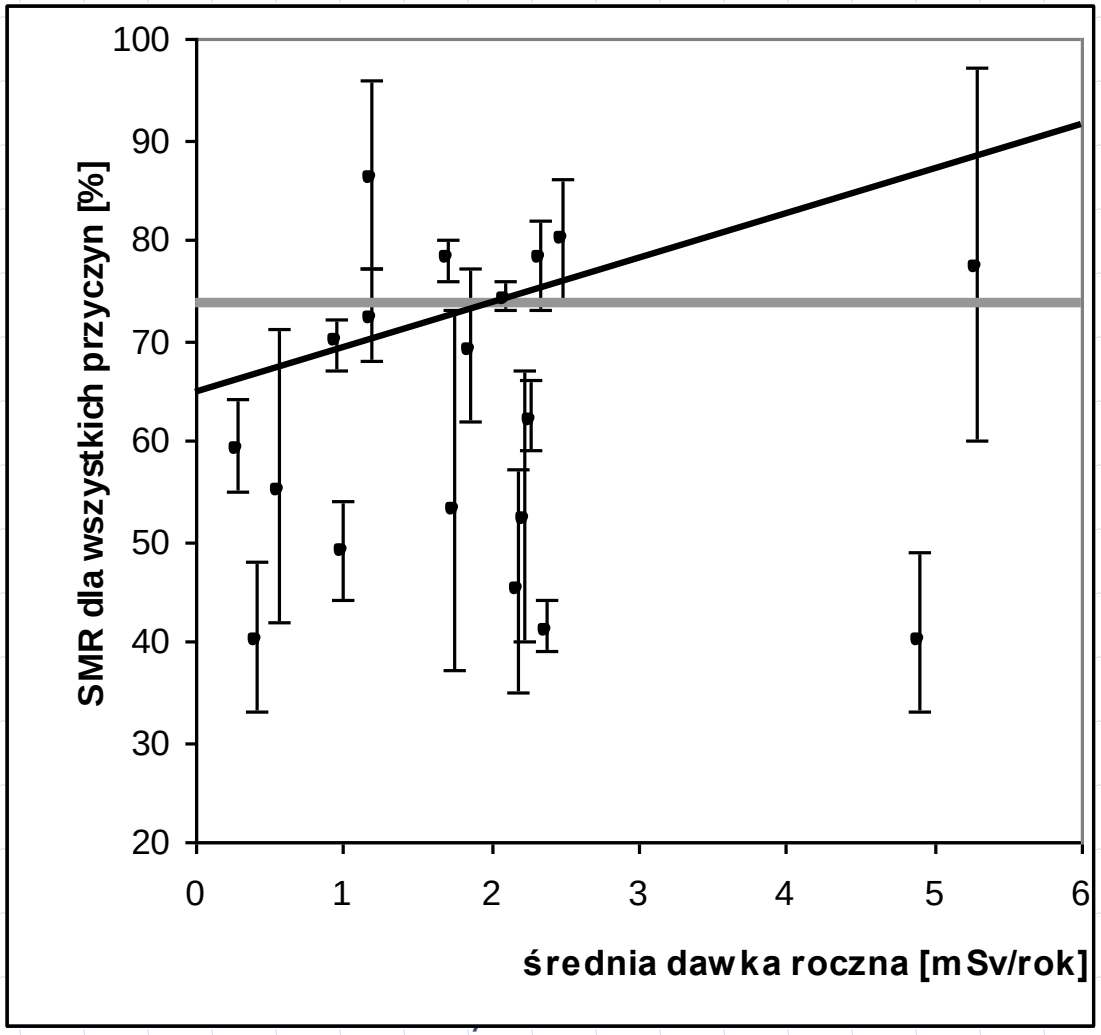








Dla dużego rozrzutu danych



Bayesowska procedura wyboru najbardziej wiarygodnego modelu

Współczynnik wyboru modelu: $W_m = \frac{P(A \mid D, I)}{P(B \mid D, I)}$

gdzie *A* oraz *B* oznaczają dwa porównywane modele, *D* dane do których się one odnoszą, a *I* wszelkie wcześniejsze informacje. Jeśli wartość W_m będzie większa od jedności, wówczas model *A* jest bardziej prawdopodobny.

Prawdopodobieństwo słuszności modelu *A* dla danych *D* oraz modelu *B* dla danych *D* można zapisać w myśl twierdzenia Bayesa jako:

$$\begin{cases} P(A|D,I) = P(D|A,I) \times P(A|I) / P(D|I) \\ P(B|D,I) = P(D|B,I) \times P(B|I) / P(D|I) \end{cases}$$

Wybór modelu – c.d.

W sytuacji, jeśli żaden z modeli *A* i *B* nie jest z góry preferowany, człony *P(A/I)* oraz *P(B/I)* są sobie równe i po podstawieniu ulegają skróceniu. Analogicznie skracają się identyczne człony *P(D/I)*.

Dla modelu *B*, zawierającego jeden parametr dopasowania λ, używając procedury marginalizacji względem tego parametru, człon *P(D/B,I)* zapisać można bayesowsko jako

$$P(D \mid B, I) = \int P(D, \lambda \mid B, I) d\lambda = \int P(D \mid \lambda, B, I) \cdot P(\lambda \mid B, I) d\lambda$$

Obierając za prior rozkład jednostajny ciągły otrzymujemy:

$$W_m = \frac{P(A \mid D, I)}{P(B \mid D, I)} = \frac{P(A \mid I)}{P(B \mid I)} \cdot \frac{P(D \mid A, I)}{P(D \mid \lambda_0, B, I)} \cdot \frac{\lambda_{\max} - \lambda_{\min}}{\delta\lambda\sqrt{2\pi}}$$

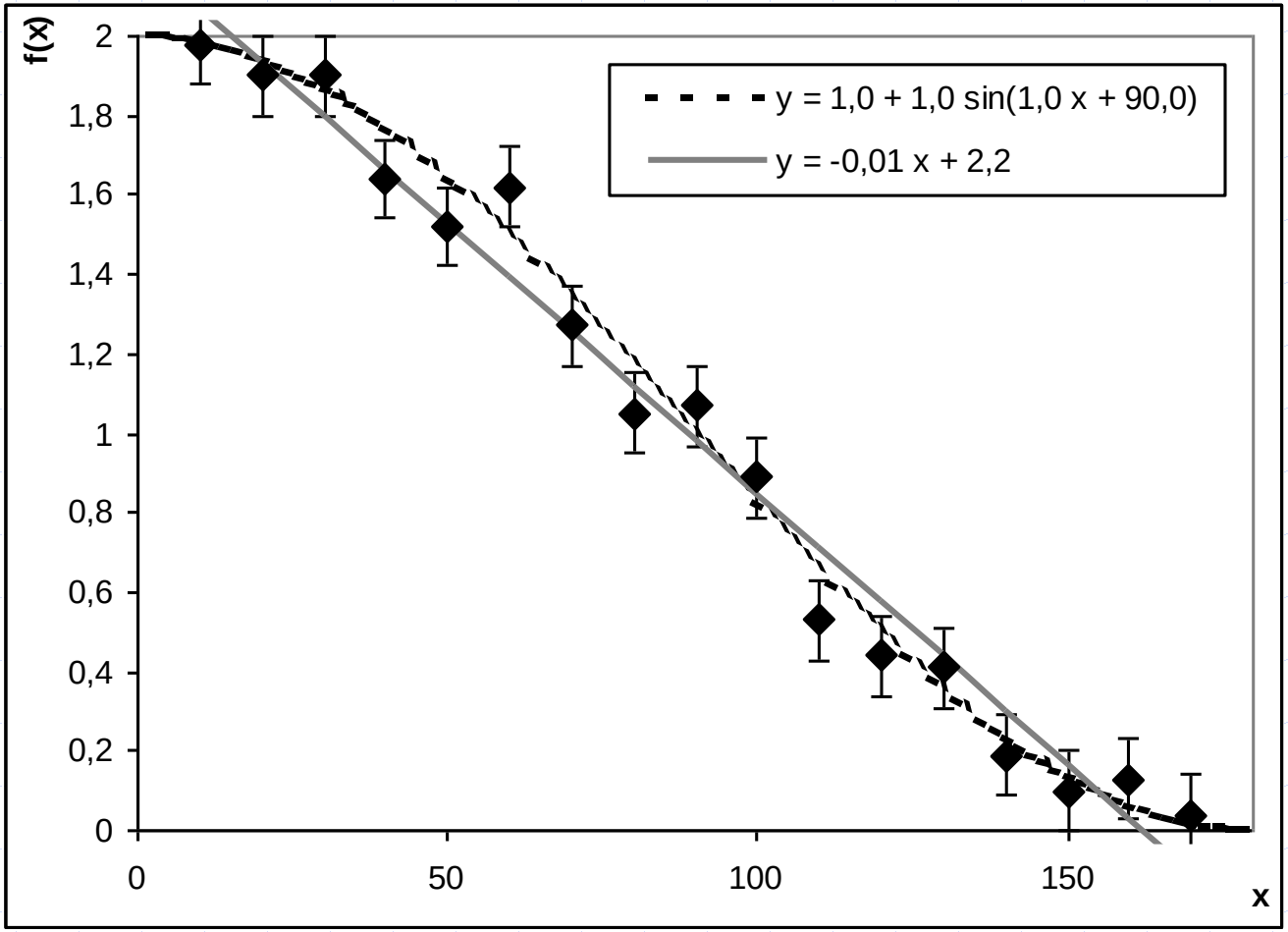
$$W_m \approx \frac{P(D \mid A, I)}{P(D \mid \lambda_0, B, I)} \cdot \frac{\lambda_{\max} - \lambda_{\min}}{\delta\lambda\sqrt{2\pi}}$$

Wybór modelu – c.d.

Dla modeli A i B zadanych wielomianami (postaci T_A i T_B o *k* i *m* parametrach) dopasowanymi do punktów eksperymentalnych E_i z niepewnościami σ_i otrzymujemy ostatecznie

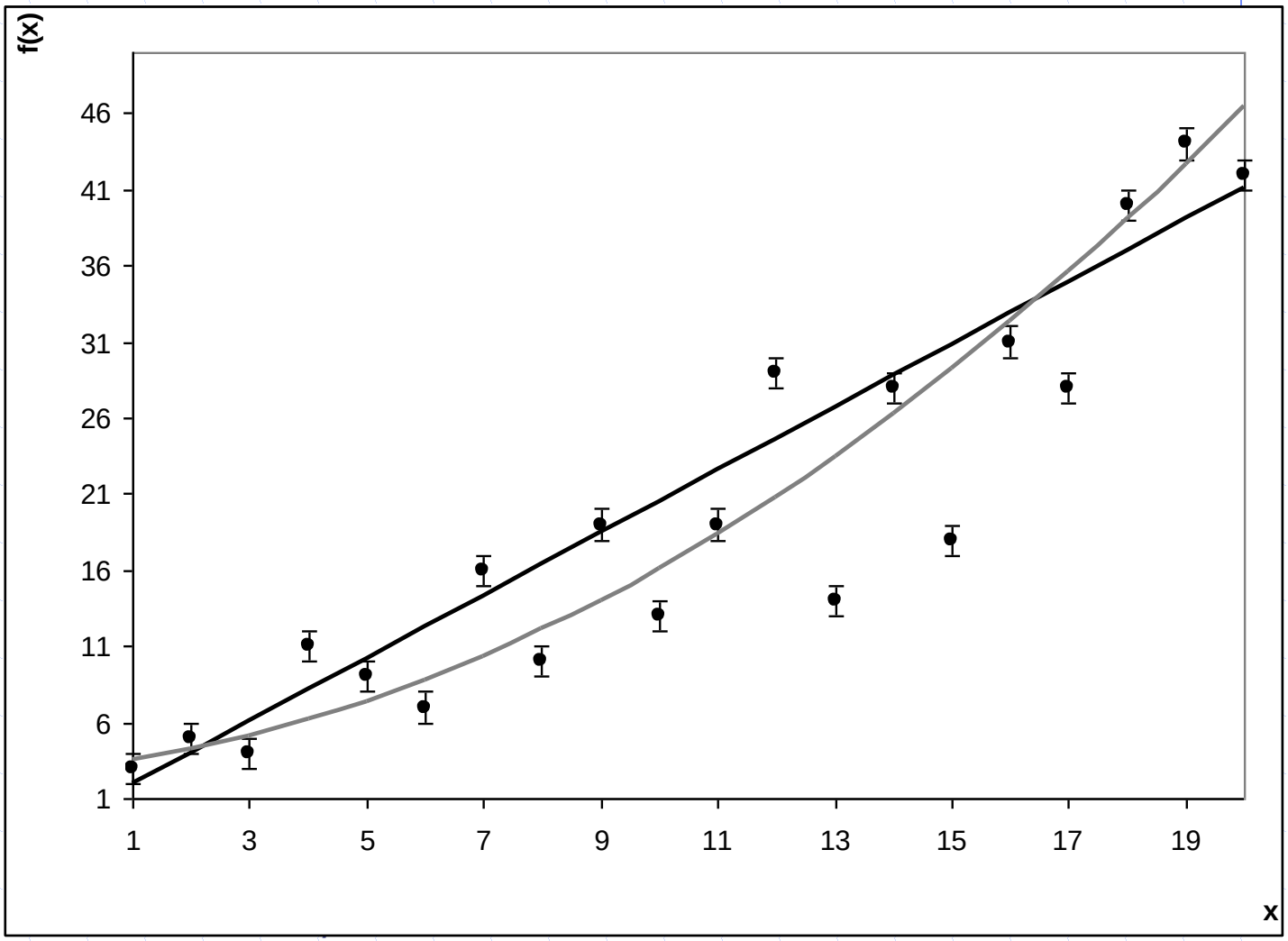
$$W_m = \frac{\sum_{i=1}^{N} \frac{1}{(T_{Ai} - E_i)^2} \left[1 - \exp\left(- \frac{(T_{Ai} - E_i)^2}{2\sigma_{0i}^2} \right) \right]}{\sum_{i=1}^{N} \frac{1}{(T_{Bi} - E_i)^2} \left[1 - \exp\left(- \frac{(T_{Bi} - E_i)^2}{2\sigma_{0i}^2} \right) \right]} \times \frac{\prod_{b=1}^{m} \frac{b_{\max}^{(B)} - b_{\min}^{(B)}}{\sigma_b^{(B)} \sqrt{2\pi}}}{\prod_{a=1}^{k} \frac{a_{\max}^{(A)} - a_{\min}^{(A)}}{\sigma_a^{(A)} \sqrt{2\pi}}}$$

Przykład zastosowania #1



Przykład zastosowania #2

- Także i tu model liniowy okazał się bardziej prawdopodobny (brzytwa Ockham'a)



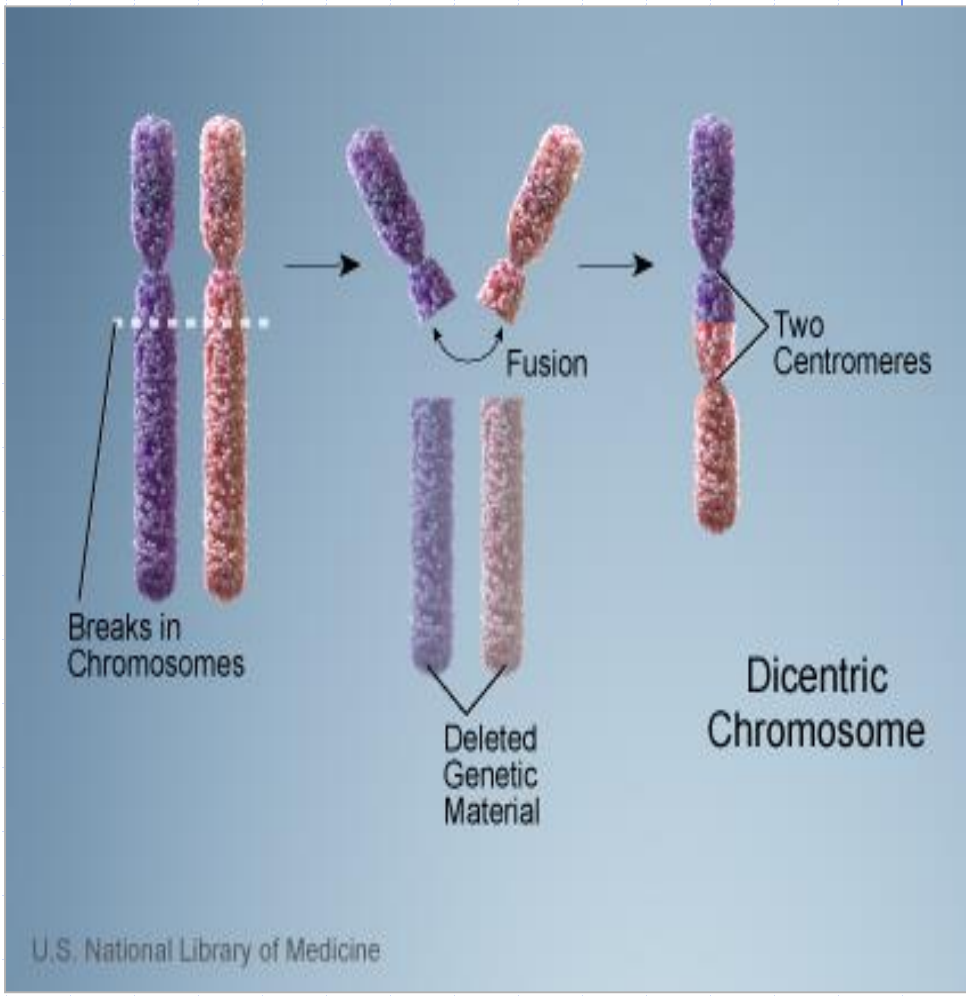
BIODOZYMETRIA CYTOGENETYCZNA DICENTRYKÓW DLA PROMIENIOWANIA MIESZANEGO

Biodozymetria cytogenetyczna dicentryków dla promieniowania mieszanego

- Biodozymetria cytogenetyczna – zespół metod/technik szacowania dawki na podstawie uszkodzeń DNA
- Dicentryki – aberracje chromosomowe dicentryczne (tj. posiadające 2 centromery)
- Promieniowanie mieszane - chodzi o napromienienie różnymi typami promieniowania jonizującego jednocześnie – w naszym przypadku będzie to promieniowanie gamma oraz neutronowe (występuje w awariach jądrowych!)

Dicentryczne aberracje chromosomowe (dicentryki)

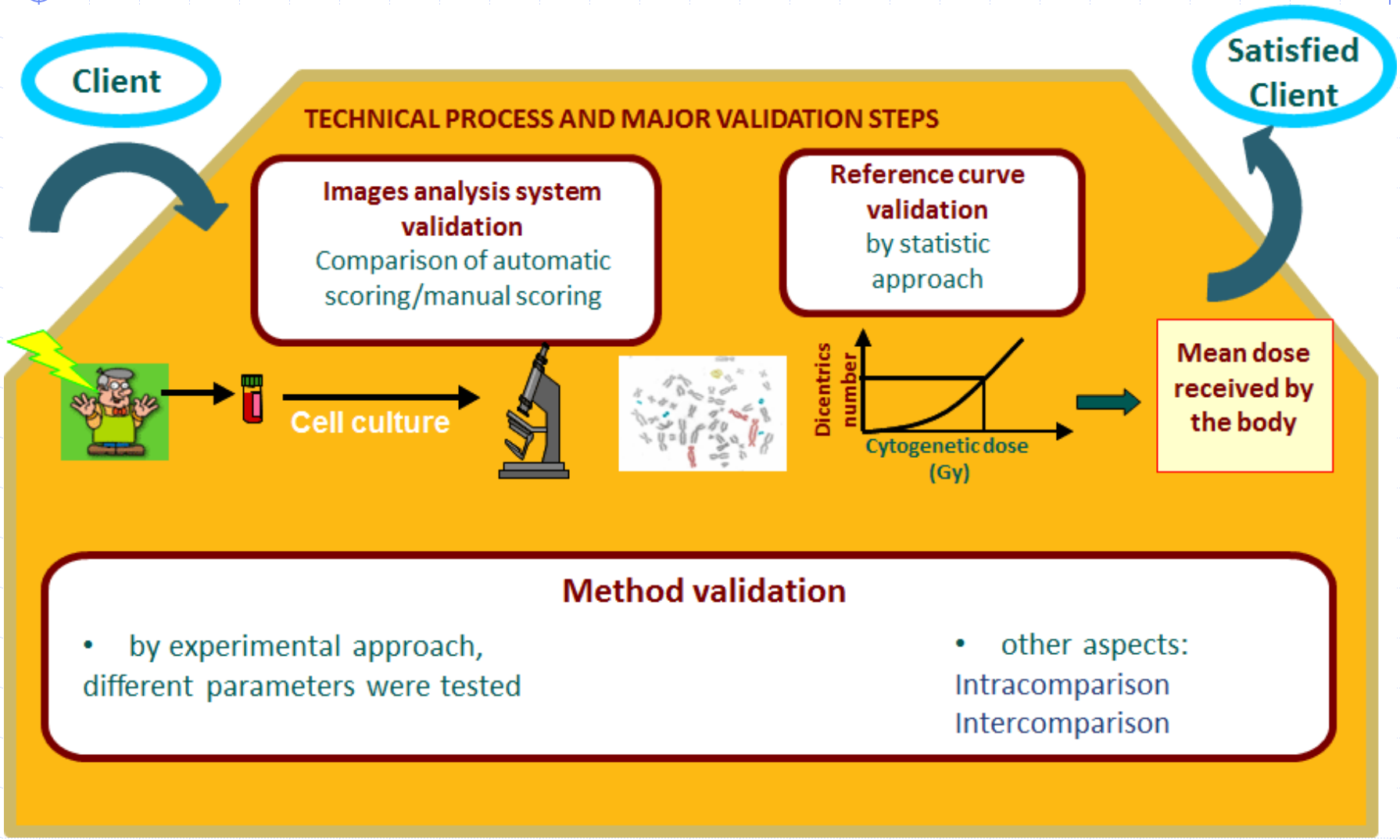
- Jedna z najgroźniejszych aberracji – chromosom dicentryczny i fragment acentryczny
- Powstaje w wyniku połączenia się dwóch chromosomów
- Aberracja nietrwała
- Prowadzi do śmierci komórki
- Fragment acentryczny, który nie zawiera centromeru w czasie mitozy nie jest segregowany do jądra komórkowego. Fragment taki obserwuje się najczęściej jako mikrojądro



Biodozymetria na dicentrykach – c.d.

- Jest to w chwili obecnej najpowszechniej używana technika w Polsce
- CLOR posiada akredytację na przeprowadzanie tego rodzaju badań
- W praktyce jest wiele metod obliczeniowych: iteracyjna, analityczna, bayesowska, Monte Carlo

Procedura w uproszczeniu



Zalety limfocytów krwi obwodowej jako modelu komórkowego do analizy dicentryków

- Łatwość pobierania
- Możliwość przechowywania i transportowania
- Zdolność do dzielenia się w hodowli po stymulacji mitogenem
- Cyrkulacja pomiędzy krwią obwodową i tkankami
- Łatwe do rozpoznania
- Powtarzalność i zgodność wyników in vivo i in vitro

Slajd dzięki uprzejmości Iwony Słoneckiej

Cytogenetyczna rekonstrukcja dawki

Częstość występowania dicentryków w limfocytach krwi obwodowej osoby narażonej na działanie promieniowania przelicza się na wartość dawki pochłoniętej za pomocą współczynników opracowanej in vitro krzywej wzorcowej (kalibracyjnej).

Fizyczna jakość promieniowania użytego do opracowania krzywej wzorcowej powinna być jak najbardziej zbliżona do promieniowania, które jest przedmiotem oceny.

Innymi słowy: w warunkach laboratoryjnych zadajemy znaną dawkę znanym typem promieniowania i patrzymy jaka liczba dicentryków jej odpowiada (in-vitro). To nam wyznacza krzywą kalibracyjną, dzięki której później szacujemy nieznaną dawkę po zliczeniu częstości występowania dicentryków w rzeczywistym przypadku napromienionego człowieka.

- Mała częstość spontaniczna, ok. 1-5 dicentryków na 1000 komórek

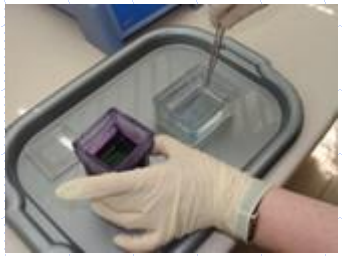
Slajd dzięki uprzejmości Iwony Słoneckiej

Cytogenetyczna rekonstrukcja dawki



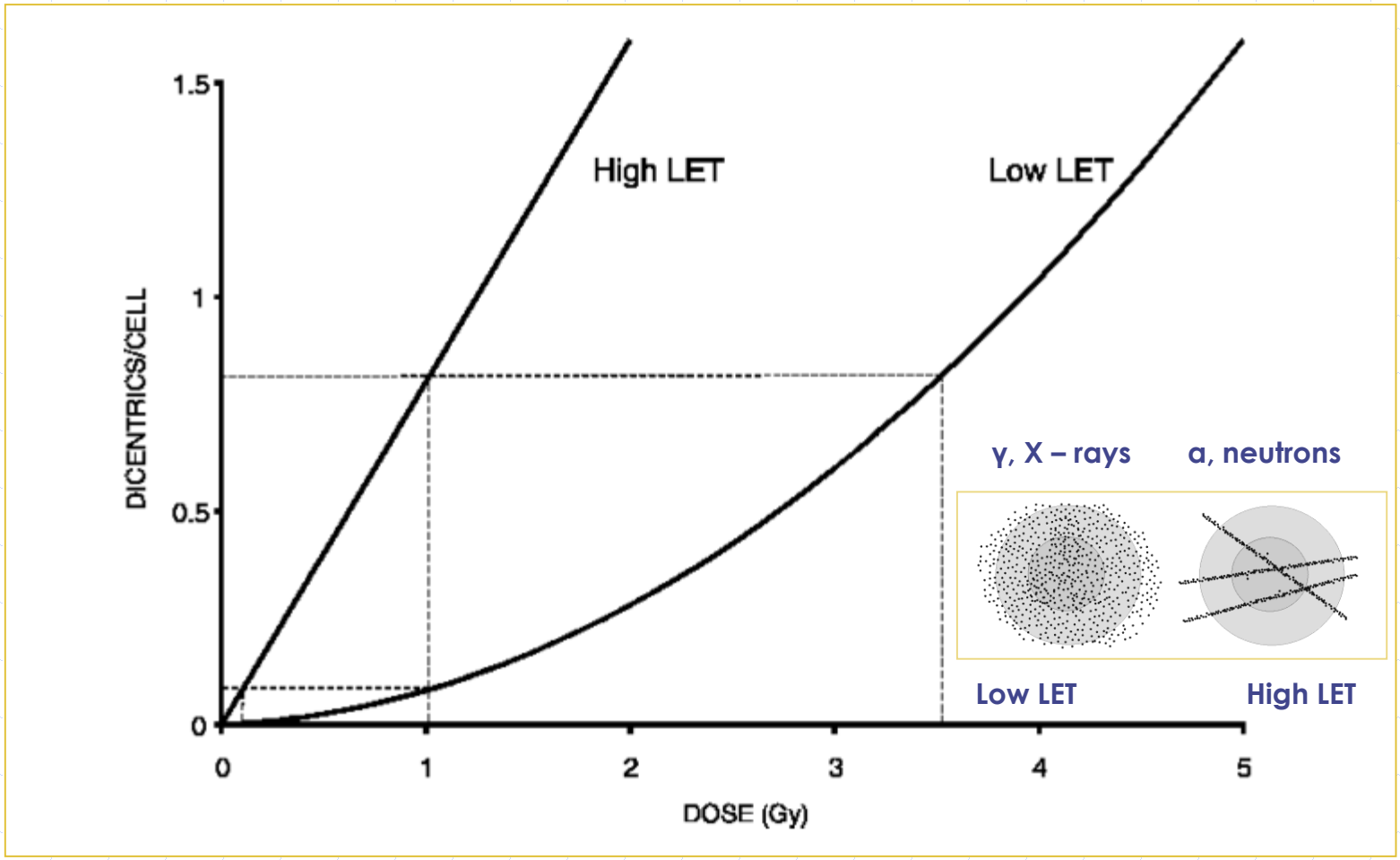






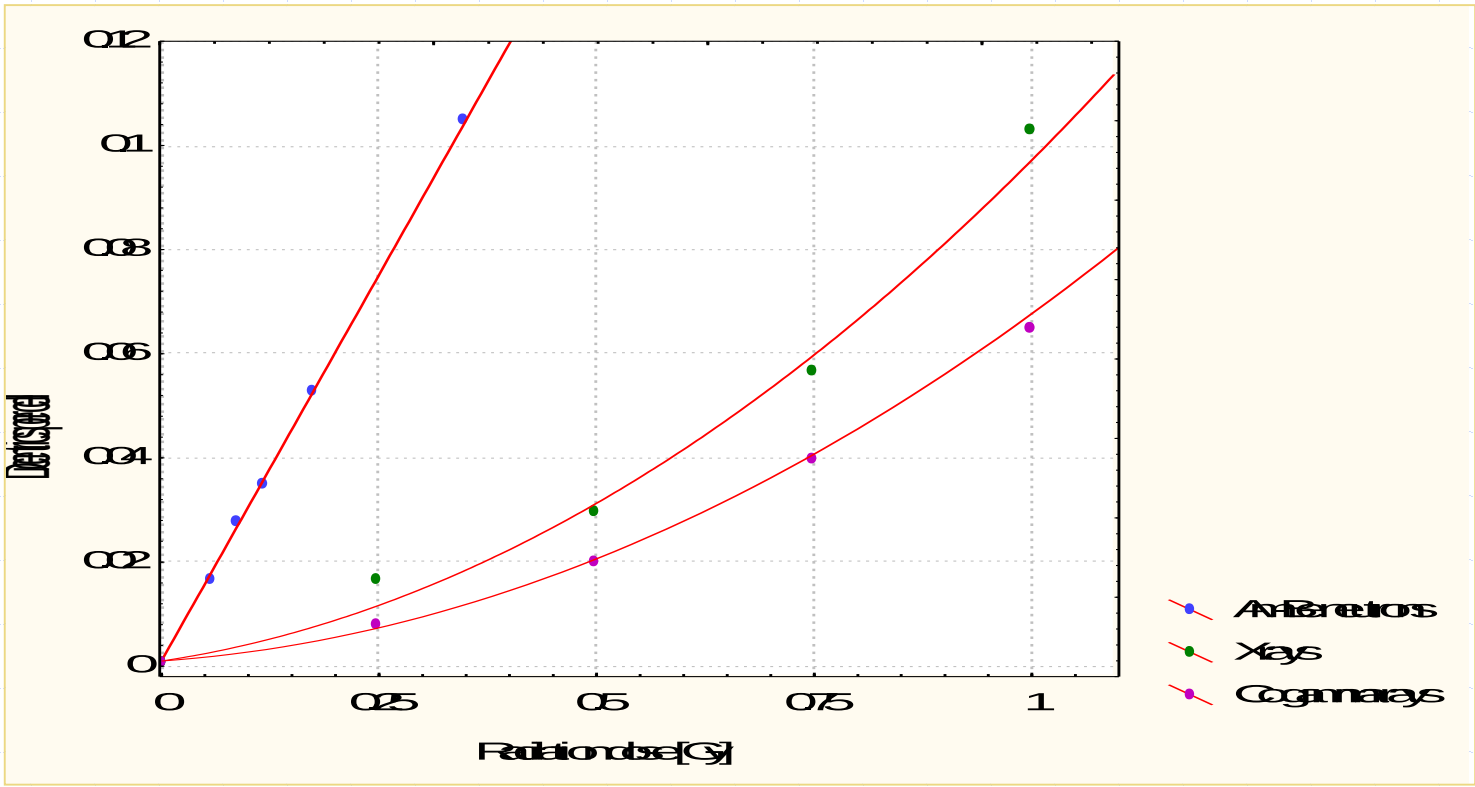

Slajd dzięki uprzejmości Iwony Słoneckiej

Krzywa kalibracyjna dla dicentryków



$$Y = C + \alpha D + \beta D^2$$

$$\boldsymbol{Y = C + \alpha D}$$



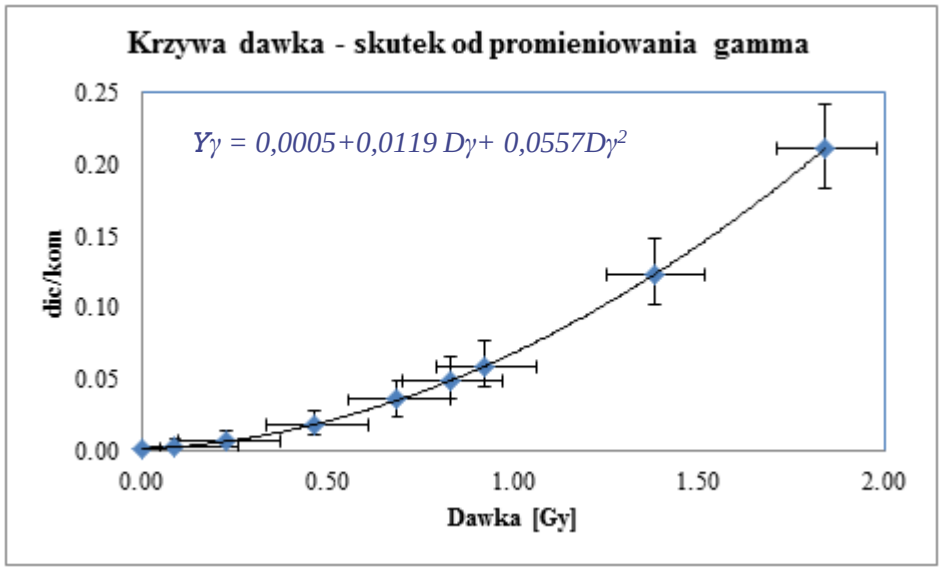
Y is the yield of dicentrics, *D* is the dose, *C* is the control (background frequency), *a* is the linear coefficient, *β* is the dose squared coefficient.

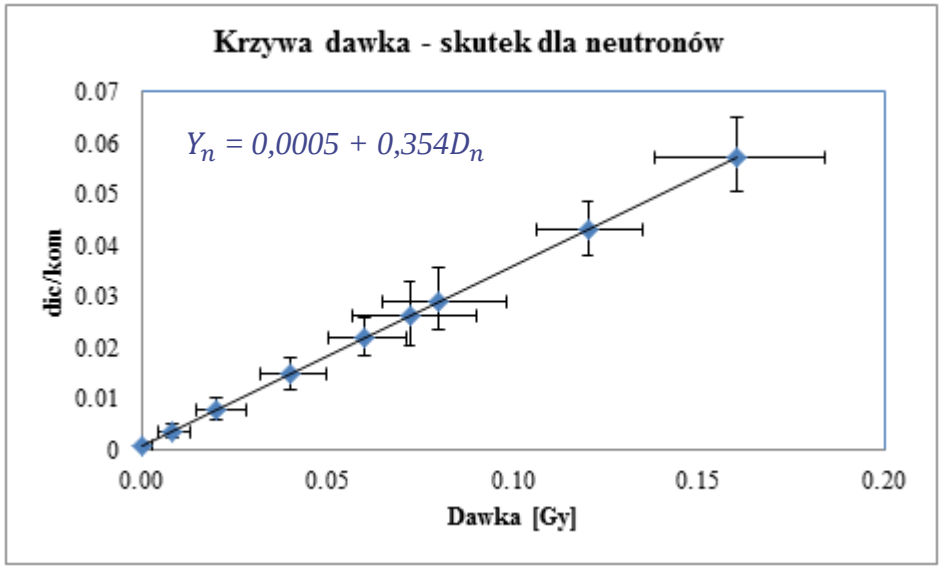
Slajd dzięki uprzejmości Iwony Słoneckiej

Krzywe kalibracyjne w CLOR dla prom. mieszanego

prom. γ $Y = C + \alpha D + \beta D^2$

prom. neutronowe $\boldsymbol{Y = C + \alpha D}$

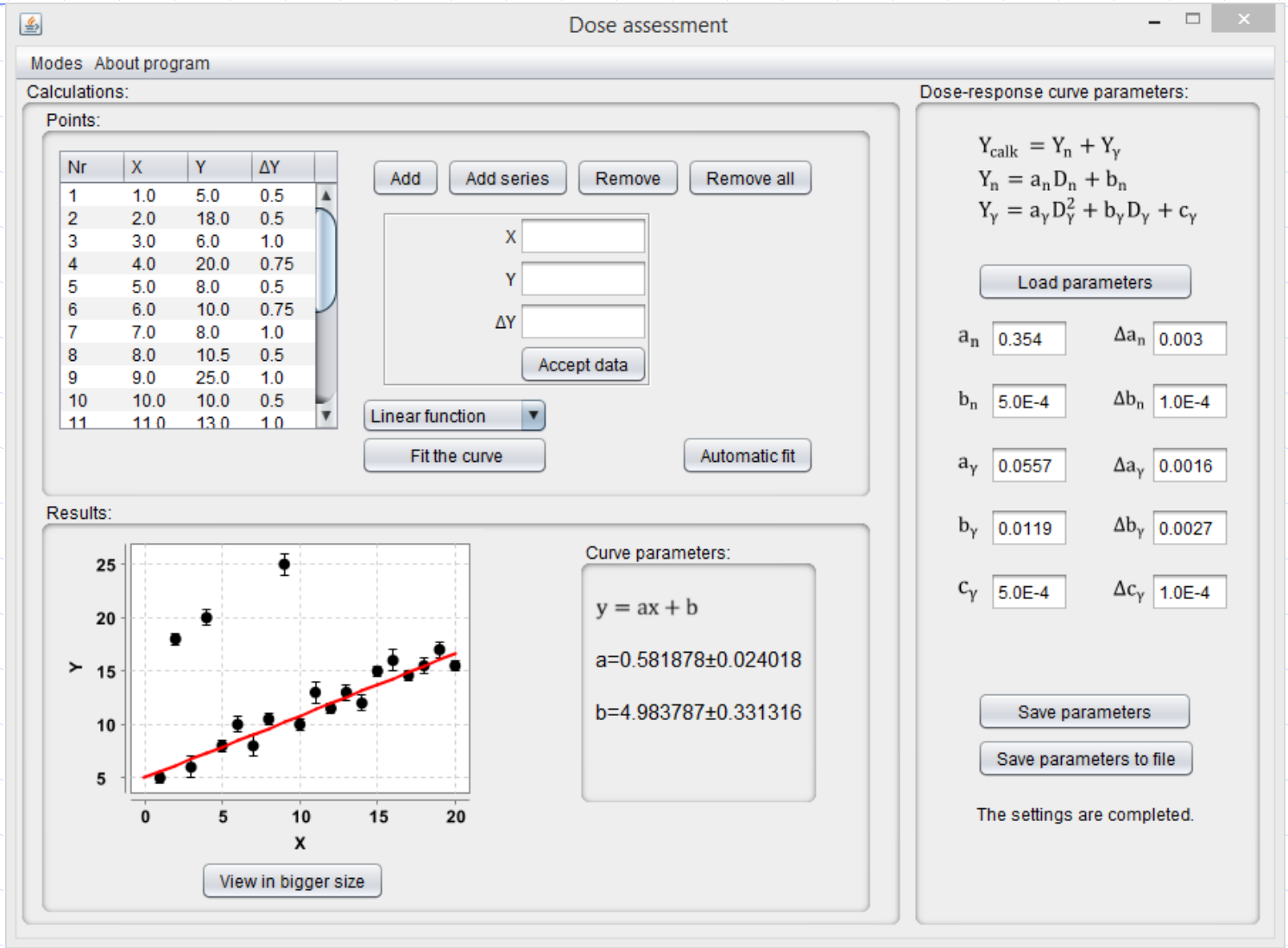




Y – częstość dicentryków,
D – dawka,
C – parametr kontrolny,
a – współczynnik liniowy,
β – współczynnik kwadratowy.

Slajd dzięki uprzejmości Iwony Słoneckiej

Krzywe można wyznaczać metodami klasycznymi lub bayesowskimi (gdy są outlier'y)



Jak szacować dawkę?

- Mając już wcześniej wyznaczoną in-vitro krzywą kalibracyjną i mikroskopowo zliczone dicentryki możemy rozpocząć szacowanie dawki
- Problem: dicentryki po fotonach niczym się nie różnią od dicentryków po neutronach
- Zazwyczaj znamy stosunek strumienia gamma/neutrony → metody klasyczne (iteracyjna i analityczna)
- Ale gdy nie znamy, to pojawia nam się dodatkowy problem → zakładamy pewien rozkład prawdopodobieństwa stosunku gamma/neutrony jako nasz prior → metody bayesowskie, metoda Monte Carlo

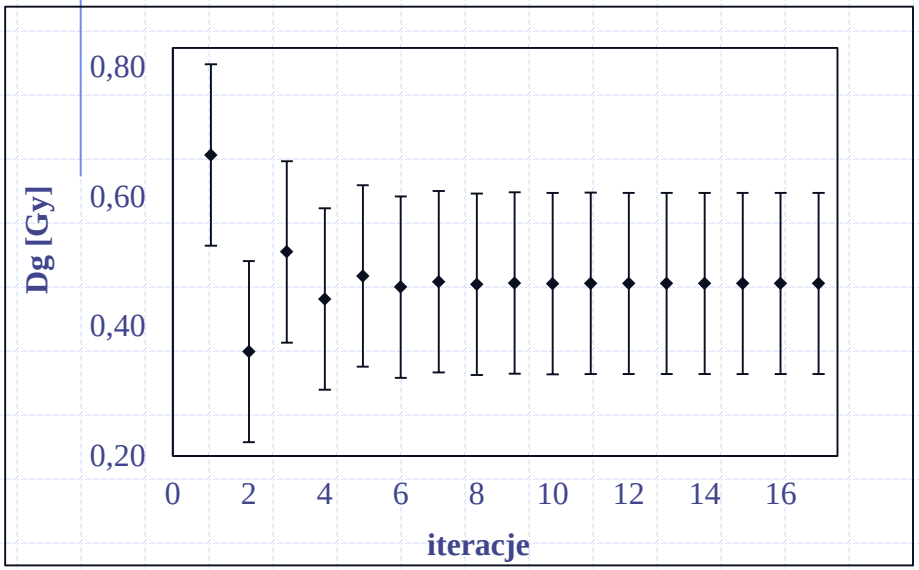
Metoda iteracyjna

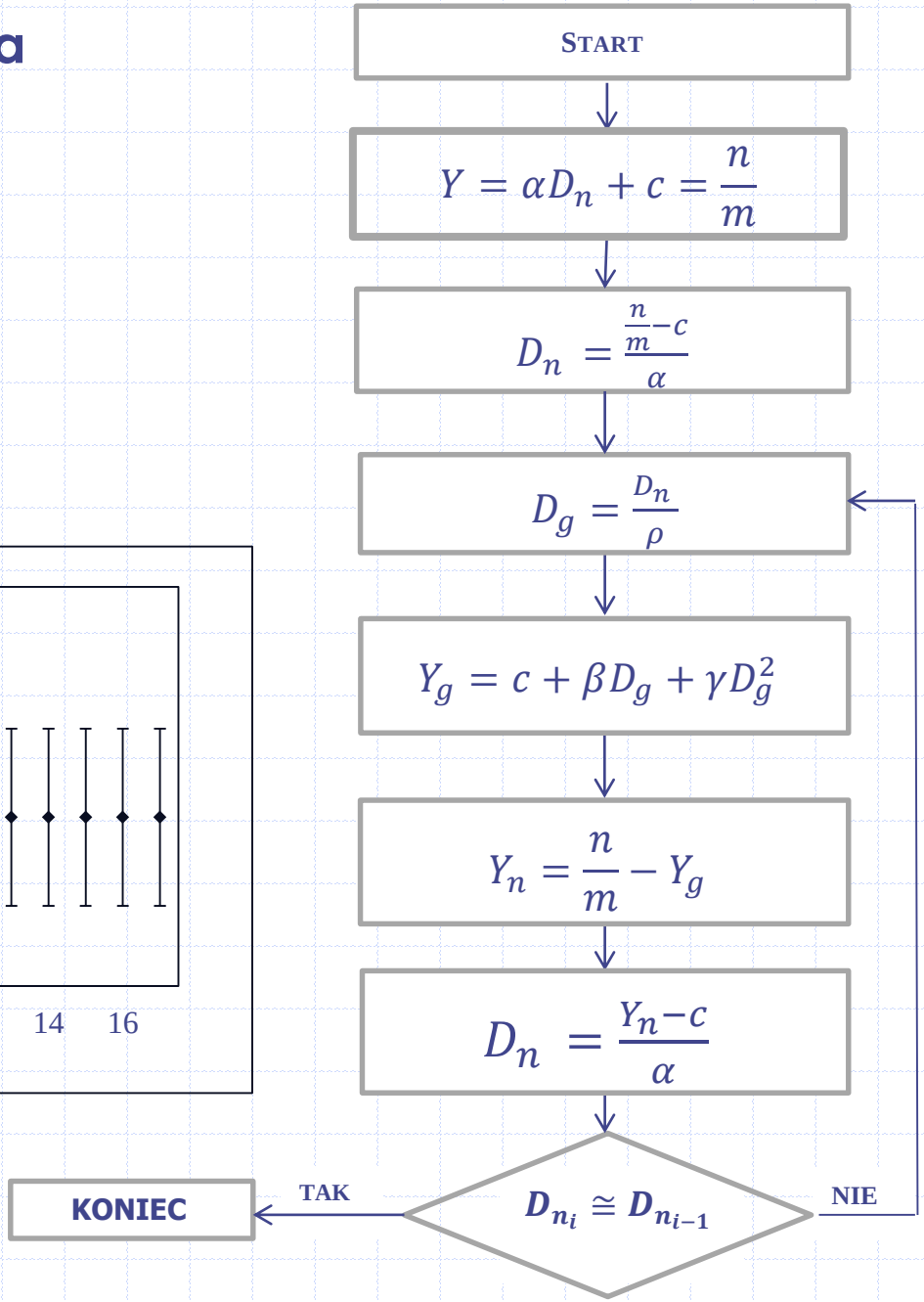
$$Y_{n+\gamma} = Y_n + Y_\gamma$$

$$\rho = D_n / D_\gamma$$

$$Y_\gamma = c + \alpha D + \beta D^2$$

$$Y_n = c + \alpha D$$





Metoda iteracyjna i analityczna

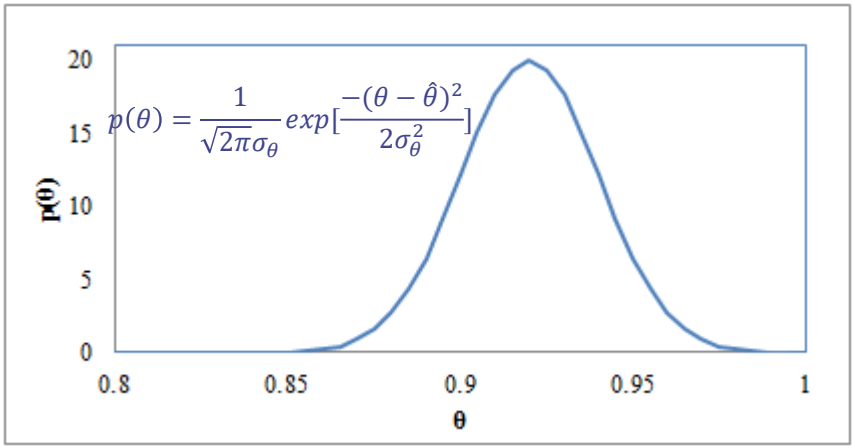
- Metoda iteracyjna jest prosta, ale poprzez efekt motyla przy kolejnych przybliżeniach zwiększa się nam niepewność
- De facto metoda ta bazuje na 2 równaniach z 2 niewiadomymi, więc wystarczy je rozwiązać (→ metoda analityczna)

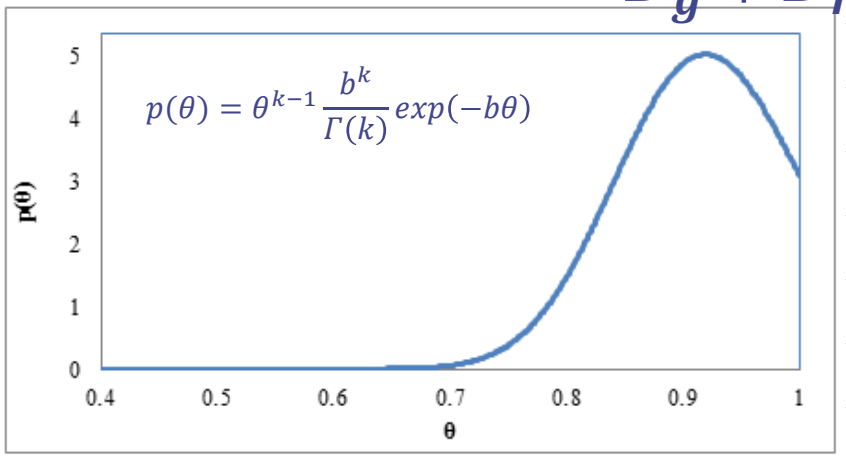
Metody bayesowskie

- Metoda iteracyjna i analityczna zakładają, że znamy stosunek strumienia fotonów do neutronów (dzięki klasycznej dozymetrii fizycznej)
- Ale dla przypadkowego narażenia osoby nieposiadającej dozymetru nie mamy tej informacji
- Wówczas stosunek ten opisujemy za pomocą prioru

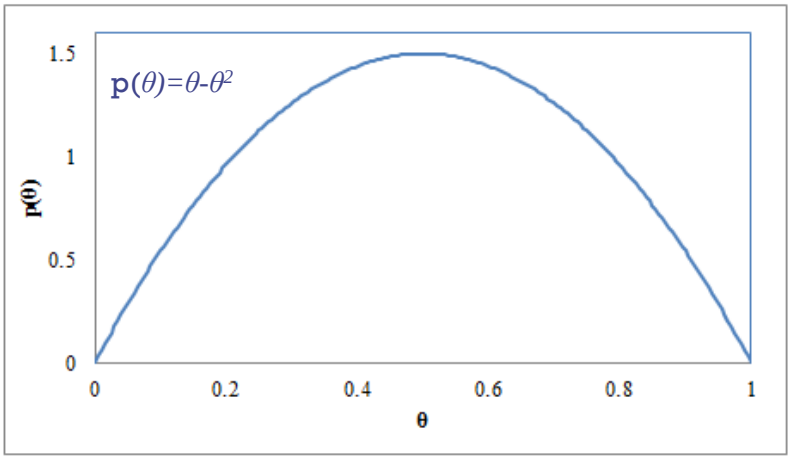
- Informatywny prior *p(θ)*:

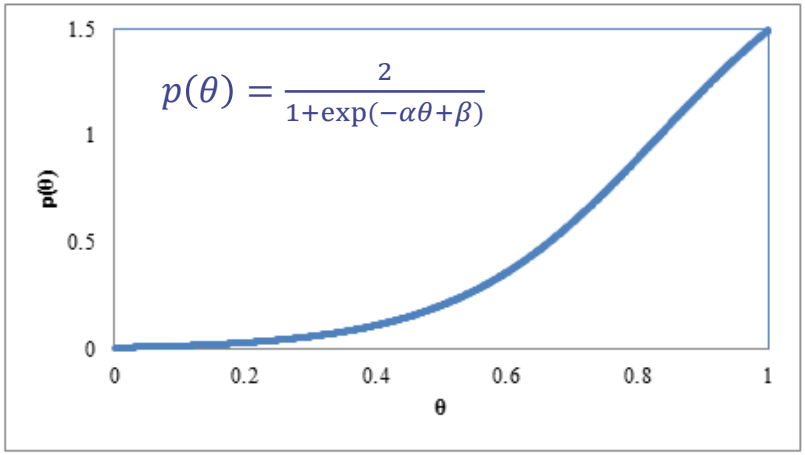
$$\boldsymbol{\theta = \frac{D_g}{D_g + D_n}}$$





- Nieinformatywny prior *p(θ)*:



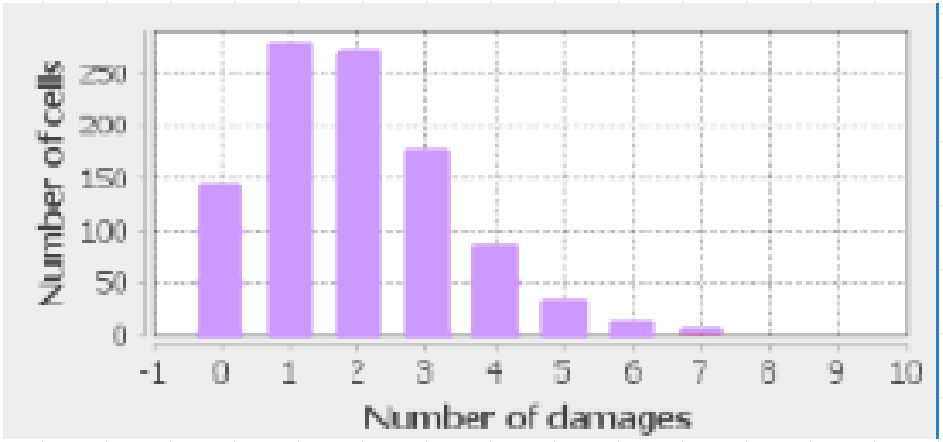


Bayesowska funkcja wiarygodności (statystyka Poissona)

$$L = \frac{(m y_f)^n \, e^{-m y_f}}{n!}$$

$$\rho = \frac{D_n}{D_g}$$

$$\theta = \frac{D_g}{D_g + D_n} = \frac{1}{\rho + 1}$$

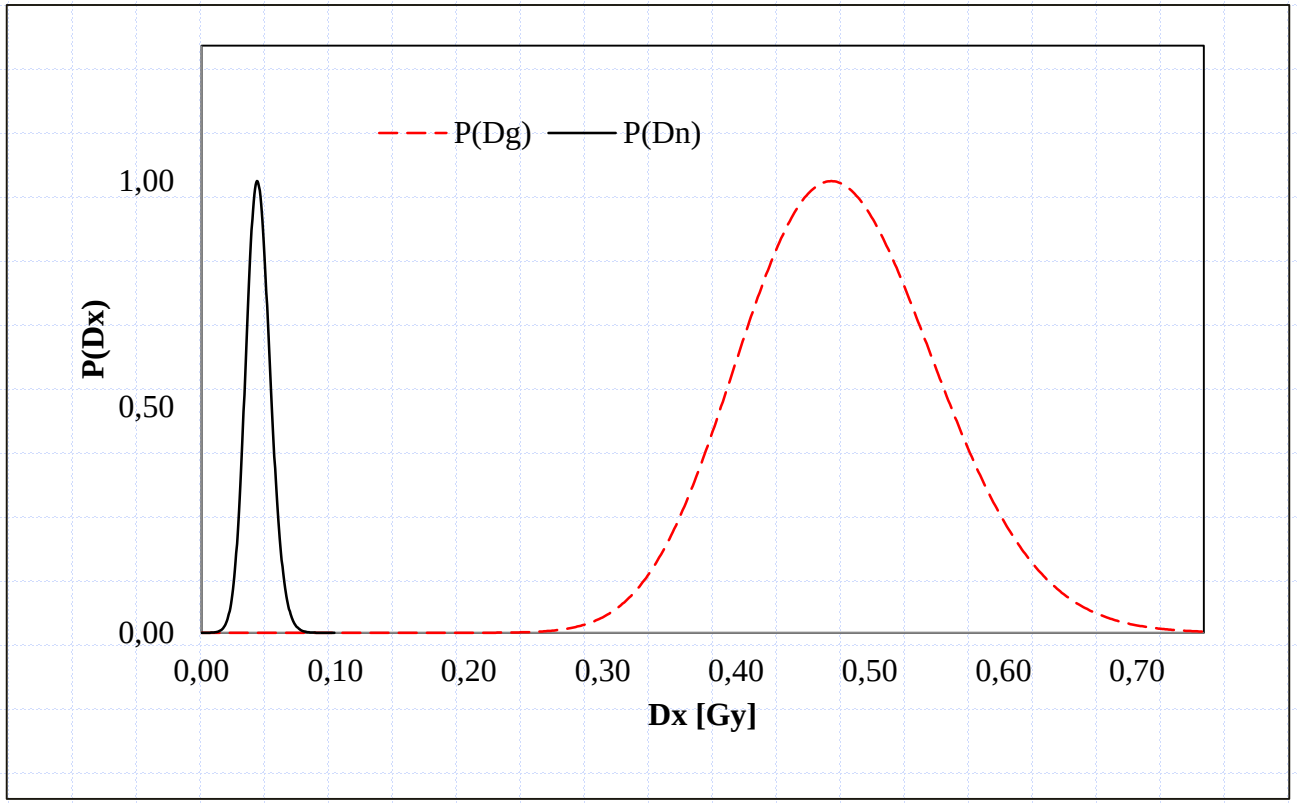


$$L(D_g|\theta) = \frac{[m\left(c + \alpha\frac{1-\theta}{\theta}D_g + \beta D_g + \gamma D_g^2\right)]^n}{n!} \cdot e^{-m(c+\alpha\frac{1-\theta}{\theta}D_g+\beta D_g+\gamma D_g^2)}$$

$$L(D_n|\theta) = \frac{[m\left(c + \alpha D_n + \beta\frac{\theta}{1-\theta}D_n + \gamma(\frac{\theta}{1-\theta}D_n\right)^2)]^n}{n!} \cdot e^{-m\left(c+\alpha D_n+\beta\frac{\theta}{1-\theta}D_n+\gamma(\frac{\theta}{1-\theta}D_n\right)^2)}$$

Dzięki temu mamy wynikowe prawdopodobieństwo a-posteriori

$$P(D_x|\theta) = \int_0^1 L(D_x|\theta)\, p(\theta) d\theta$$



Dose assessment

Modes About program

Iterative method | Analytical method | Bayesian method

Calculations:

Prior:

Gaussian

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma_\theta} \exp\left[\frac{-(\theta-\hat{\theta})^2}{2\sigma_\theta^2}\right]$$

$\hat{\theta}$ 0.92

σ_θ 0.92

Draw prior

Sample parameters:

Number of dicentrics:

Number of cells:

Calculate

Results:

D_n [Gy]

D_γ [Gy]

Y_n [dic/cell]

Y_γ [dic/cell]

Dose-response curve parameters:

$$Y_{calk} = Y_n + Y_\gamma$$
$$Y_n = a_n D_n + b_n$$
$$Y_\gamma = a_\gamma D_\gamma^2 + b_\gamma D_\gamma + c_\gamma$$

Load parameters

a_n	0.354	Δa_n	0.003
b_n	5.0E-4	Δb_n	1.0E-4
a_γ	0.0557	Δa_γ	0.0016
b_γ	0.0119	Δb_γ	0.0027
c_γ	5.0E-4	Δc_γ	1.0E-4

Save parameters

Save parameters to file

The settings are completed.

Absorbed dose assessment

Initial data

Measured chromosome aberration frequency: y_f

Calibration curves:

$Y_n = \alpha D_n + c$ — α | β | γ | c

$Y_g = \beta D_g + \gamma D_g^2 + c$ — $\Delta\alpha$ | $\Delta\beta$ | $\Delta\gamma$ | Δc

Open from file

Save to file

Clear

Use a prior probability distribution

Gaussian

$$p(\theta) \propto \frac{1}{\sqrt{2\pi}\sigma_\theta} \exp\left[-\frac{(\theta-\hat{\theta})^2}{2\sigma_\theta^2}\right]$$

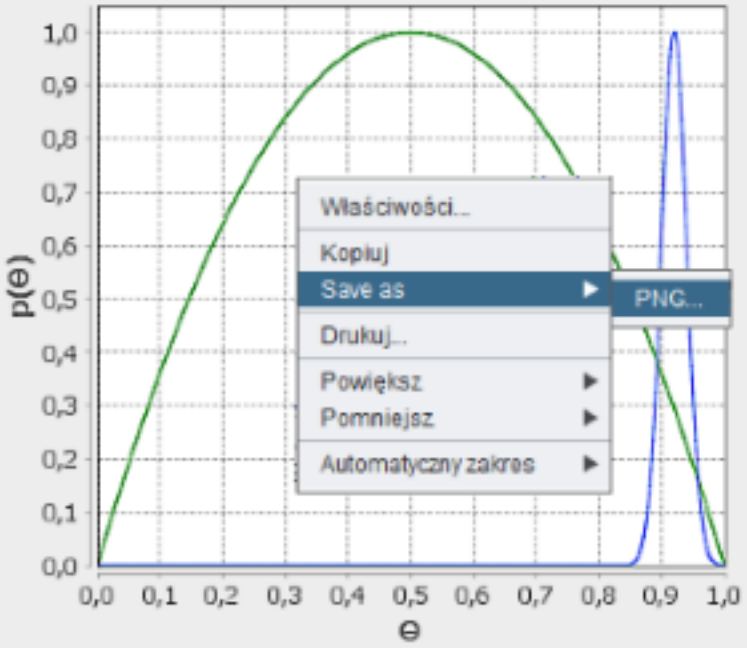
- Gaussian
- Scaled Gaussian
- Beta
- Simplified Beta
- Sigmoidal
- Avrami
- Uniform
- Constant

σ_θ 0.02

σ_θ

ρ σ_ρ

Draw prior | Clear graph

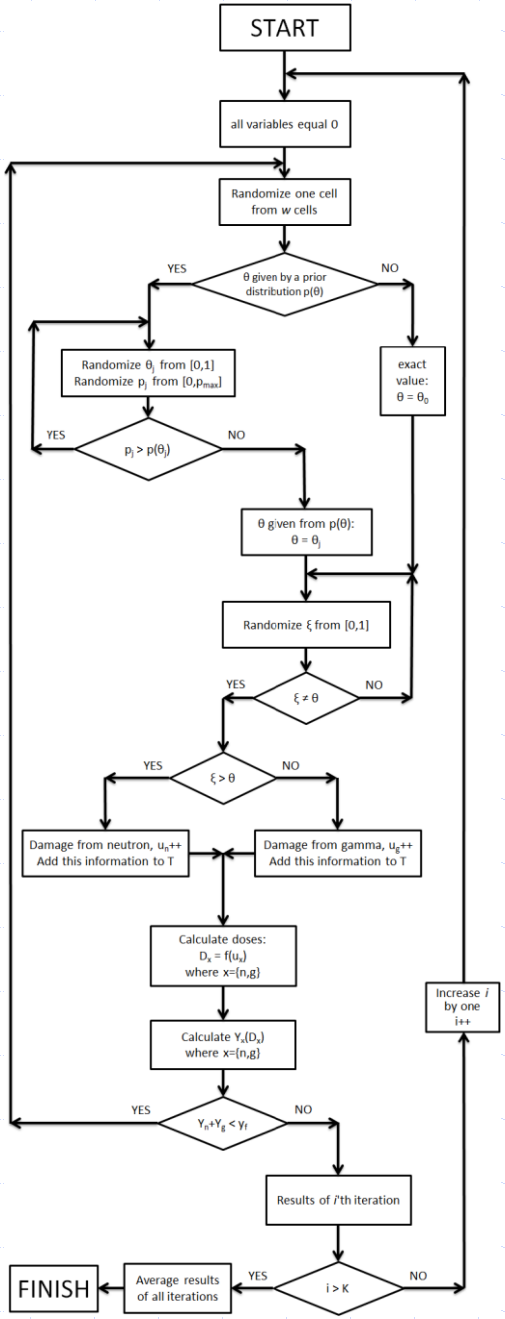


Calculate

Results

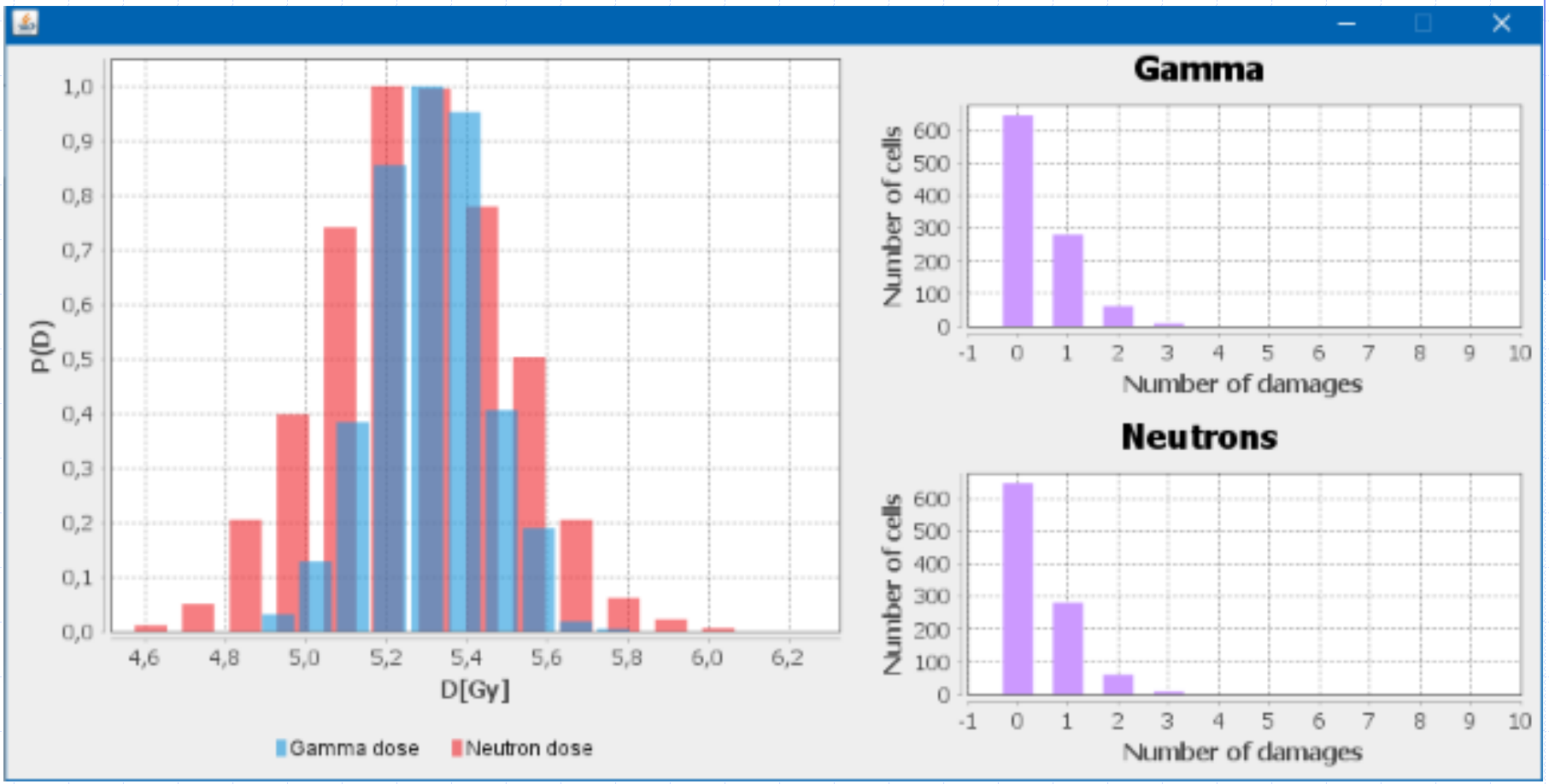
D_n [Gy] | D_g [Gy] | D_{n+g} [Gy]

Y_n [dic/cell] | Y_g [dic/cell] | Y_{n+g} [dic/cell]

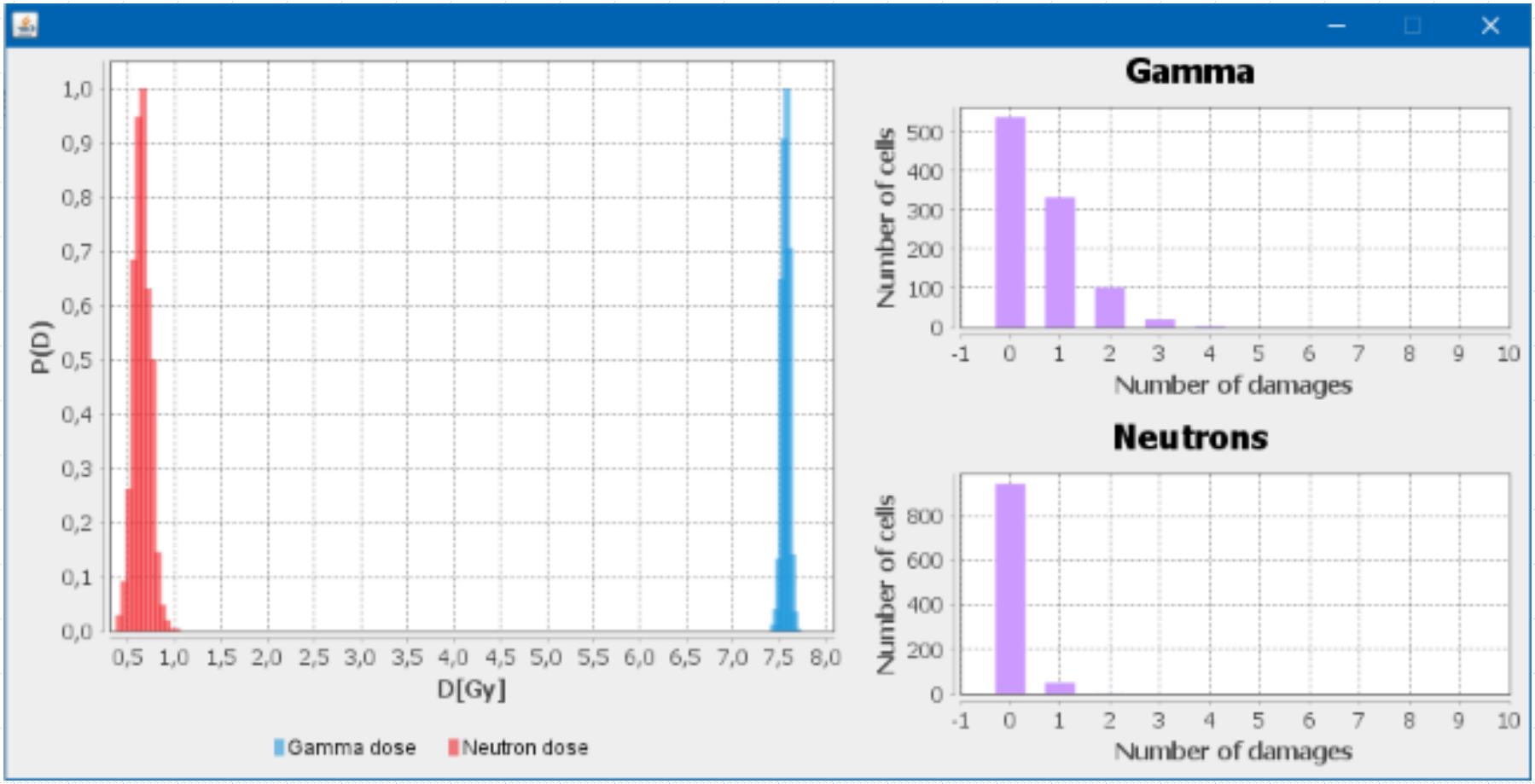


Metoda Monte Carlo

- podanie wartości θ lub priora
- losowanie komórek z próbki wirtualnej
- program wykonuje kolejne iteracje, aż do momentu osiągnięcia w próbce podanej na wstępie częstości aberracji
- obliczenie dawek pochłoniętych od neutronów i od promieniowania gamma



Results for yf =3.5 [dic/cell], θ=0.5, Dt=10.59 [Gy].



Results for yf =3.5 [dic/cell], θ=0.92, Dt=8.2 [Gy].

Cornell University Library, arXiv.org, 2014

Alternative statistical methods for cytogenetic radiation biological dosimetry

Krzysztof Wojciech Fornalski [1,2]

[1)] PGE EJ 1 Sp. z o.o., Technology and Operations Office, ul. Mysia 2, 00-496 Warszawa, Poland
[2)] contract of specified task with Central Laboratory for Radiological Protection (CLOR), ul. Konwaliowa 7, 03-194 Warszawa, Poland
e-mail: krzysztof.fornalski@gmail.com, krzysztof.fornalski@gkpge.pl

ABSTRACT

The paper presents alternative statistical methods for biological dosimetry, such as the Bayesian and Monte Carlo method. The classical Gaussian and robust Bayesian fit algorithms for the linear, linear-quadratic as well as saturated and critical calibration curves are described. The Bayesian model selection algorithm for those curves is also presented. In