

Metody matematyczne astronomii 2

Wykład 6

Mariusz Tarnopolski

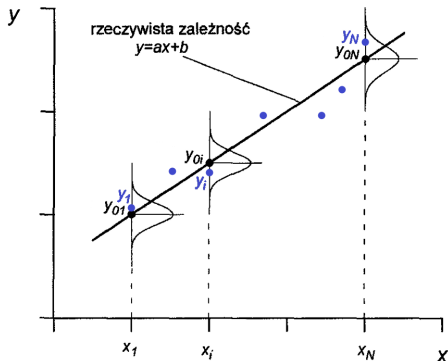
Instytut Astronomii UMK

MMA 2 ©2023



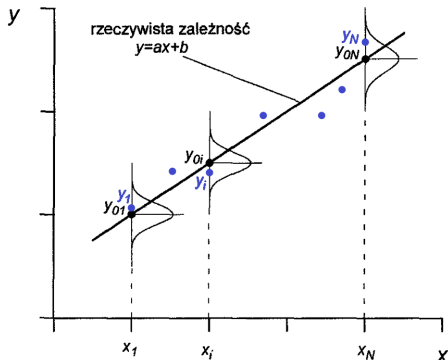
Zał.

Rozkład odchyłeń punktów pomiarowych wokół linii najlepszego dopasowania jest gaussowski, zlokalizowany na linii oraz o stałej wariancji $\sigma^2 = \sigma_i^2$ (homoskedastyczność).



Zał.

Rozkład odchyłeń punktów pomiarowych wokół linii najlepszego dopasowania jest gaussowski, zlokalizowany na linii oraz o stałej wariancji $\sigma^2 = \sigma_i^2$ (homoskedastyczność).



Chcemy dopasować funkcję $y = g(x; \vec{a})$, z parametrami $\vec{a} = (a_1, a_2, \dots, a_p)$, które maksymalizują łączne prawdopodobieństwo punktów pomiarowych (tj. tak by względem linii najlepszego dopasowania to prawd. było maksymalne):

$$L = \prod_{i=1}^N f(y_i; g(x_i), \sigma^2)$$

L — funkcja wiarygodności (*likelihood*); niech f to Gaussy:

$$L = \prod_{i=1}^N f(y_i; g(x_i), \sigma^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - g(x_i))^2}{2\sigma^2}\right)$$

L — funkcja wiarygodności (*likelihood*); niech f to Gaussy:

$$L = \prod_{i=1}^N f(y_i; g(x_i), \sigma^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - g(x_i))^2}{2\sigma^2}\right)$$

Logarytm funkcji wiarygodności (*loglikelihood*):

$$\mathcal{L} \equiv \ln L = \sum_{i=1}^N \ln f(y_i; g(x_i), \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - g(x_i))^2$$

L — funkcja wiarygodności (*likelihood*); niech f to Gaussy:

$$L = \prod_{i=1}^N f(y_i; g(x_i), \sigma^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - g(x_i))^2}{2\sigma^2}\right)$$

Logarytm funkcji wiarygodności (*loglikelihood*):

$$\mathcal{L} \equiv \ln L = \sum_{i=1}^N \ln f(y_i; g(x_i), \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - g(x_i))^2$$

Rozwiązując $\frac{\partial \mathcal{L}}{\partial a_k} = 0$, $k = 1, \dots, p$, dostaniemy takie parametry \vec{a} , które maksymalizują \mathcal{L} , czyli najlepsze dopasowanie (funkcją g).

L — funkcja wiarygodności (*likelihood*); niech f to Gaussy:

$$L = \prod_{i=1}^N f(y_i; g(x_i), \sigma^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - g(x_i))^2}{2\sigma^2}\right)$$

Logarytm funkcji wiarygodności (*loglikelihood*):

$$\mathcal{L} \equiv \ln L = \sum_{i=1}^N \ln f(y_i; g(x_i), \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - g(x_i))^2$$

Rozwiązując $\frac{\partial \mathcal{L}}{\partial a_k} = 0$, $k = 1, \dots, p$, dostaniemy takie parametry \vec{a} , które maksymalizują \mathcal{L} , czyli najlepsze dopasowanie (funkcją g). Dla regresji liniowej wychodzi to samo co metodą najmniejszych kwadratów.

Ponadto

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - g(x_i))^2 \equiv \frac{RSS}{N}$$

gdzie RSS — *residual sum of squares*.

Ponadto

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - g(x_i))^2 \equiv \frac{RSS}{N}$$

gdzie RSS — *residual sum of squares*.

Po podstawieniu $\hat{\sigma}^2$ do \mathcal{L} dostaniemy

$$\mathcal{L}_{\max} = -\frac{N}{2} (\ln 2\pi \hat{\sigma}^2 + 1) = -\frac{N}{2} (\ln 2\pi + 1) - \frac{N}{2} \ln \frac{RSS}{N}$$

gdzie pierwszy człon — dla danego zestawu danych — jest stały, zatem można go pominąć (formalnie: tak przeskalować \mathcal{L} by go usunąć).

Ponadto

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - g(x_i))^2 \equiv \frac{RSS}{N}$$

gdzie RSS — *residual sum of squares*.

Po podstawieniu $\hat{\sigma}^2$ do \mathcal{L} dostaniemy

$$\mathcal{L}_{\max} = -\frac{N}{2} (\ln 2\pi \hat{\sigma}^2 + 1) = -\frac{N}{2} (\ln 2\pi + 1) - \frac{N}{2} \ln \frac{RSS}{N}$$

gdzie pierwszy człon — dla danego zestawu danych — jest stały, zatem można go pominąć (formalnie: tak przeskalować \mathcal{L} by go usunąć).

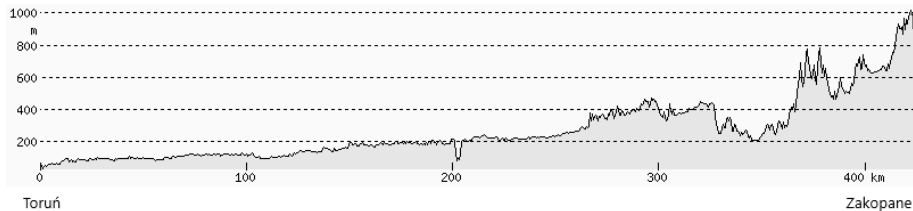
Ostatecznie dla regresji (czy to liniowej, czy nieliniowej) można:

$$\mathcal{L}_{\max} = -\frac{N}{2} \ln \frac{RSS}{N}$$

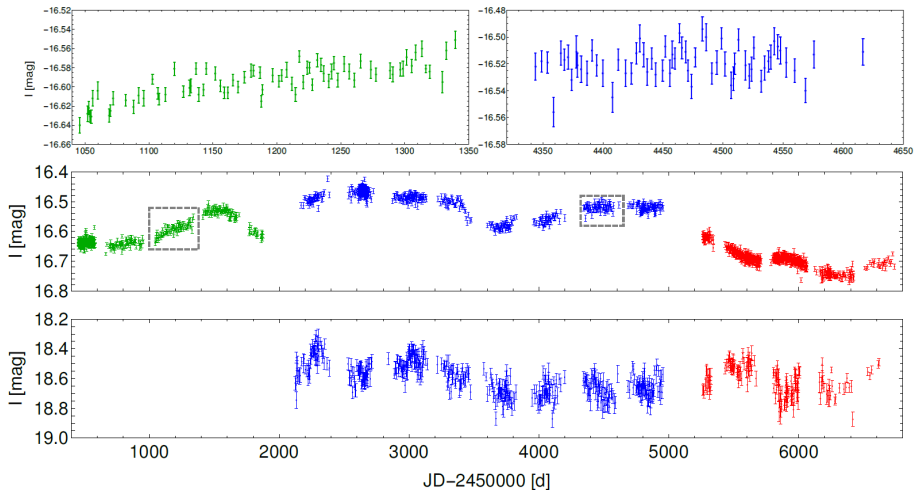
Uporządkowany ciąg wartości. Najczęściej pokazuje zmienność w czasie, jednak zmienna niezależna nie musi być czasem, może być dowolną wielkością, która porządkuje zmienne od niej zależne.

Uporządkowany ciąg wartości. Najczęściej pokazuje zmienność w czasie, jednak zmienna niezależna nie musi być czasem, może być dowolną wielkością, która porządkuje zmienne od niej zależne.

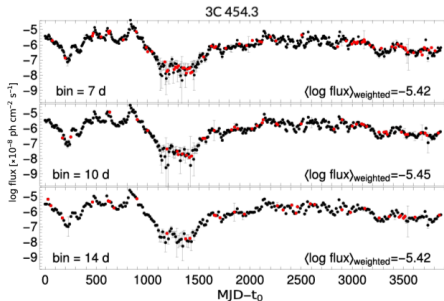
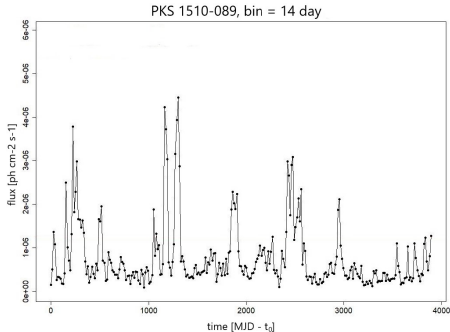
Np. wykres wzniesienia terenu też jest szeregiem czasowym:



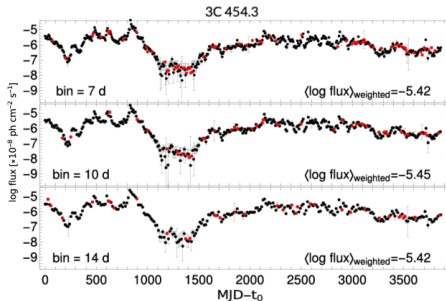
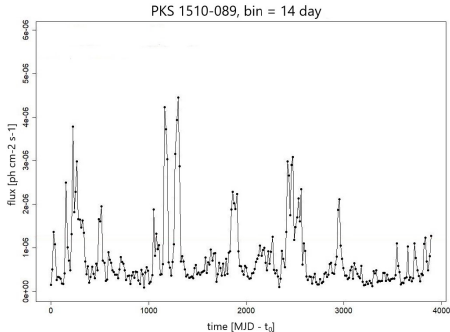
Krzywe zmian blasku kandydatów na blazary za Obłokami Magellana
(*Optical Gravitational Lensing Experiment, OGLE*):



Krzywe zmian blasku blazarów w zakresie gamma (Fermi Large Area Telescope):



Krzywe zmian blasku blazarów w zakresie gamma (Fermi Large Area Telescope):

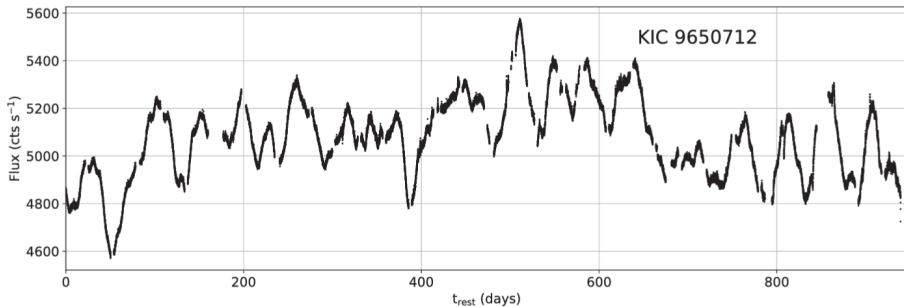


$$f(t) = \exp[l(t)]$$

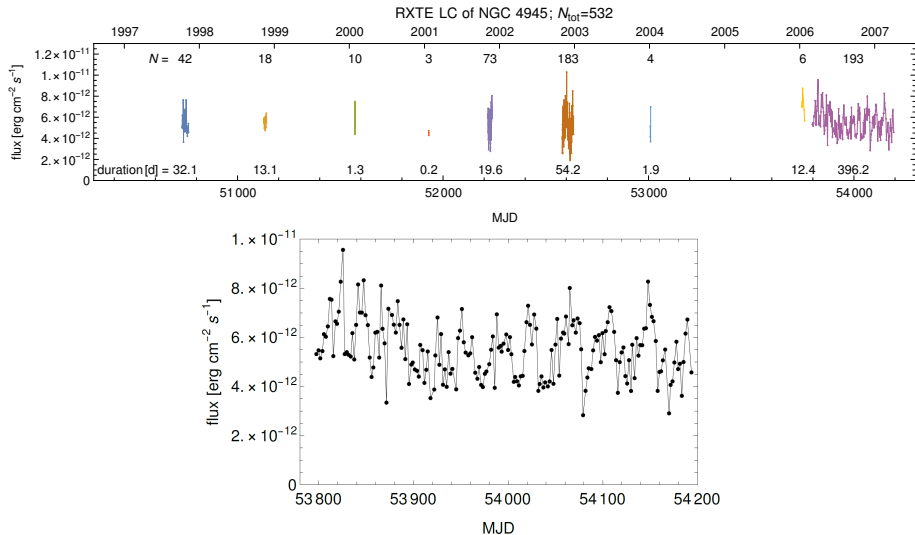
f — strumień

l — proces stochastyczny rządzący obserwowaną zmiennością

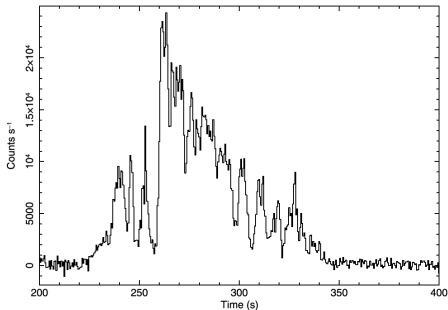
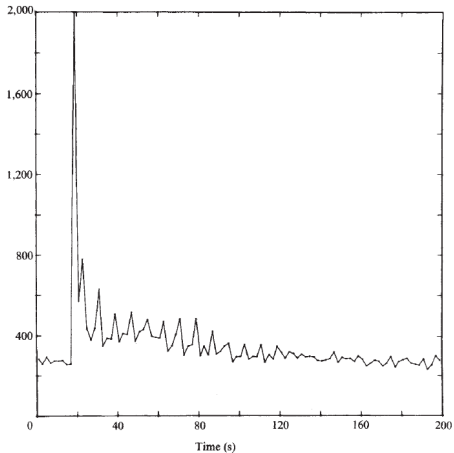
Kosmiczny Teleskop Keplera:



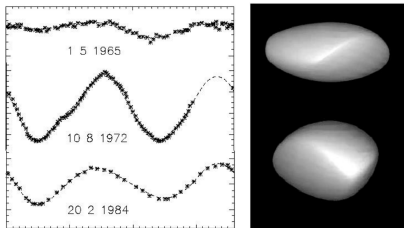
Promieniowanie X (Rossi X-ray Timing Explorer):



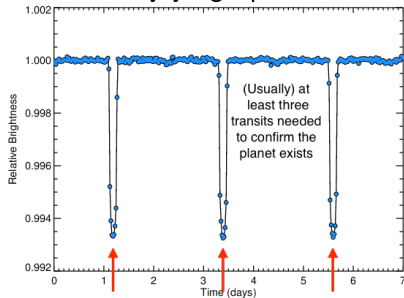
Rozbłyski gamma:



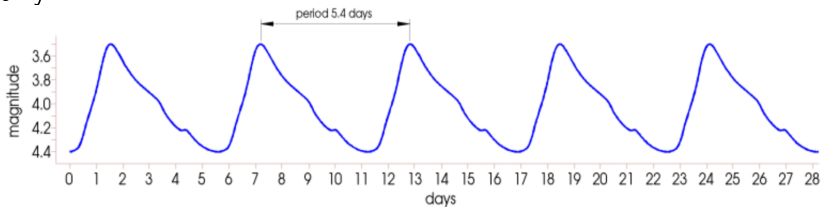
Asteroidy:



Tranzyty egzoplanet:



Cefeidy:



Light curve of Delta Cephei

© 2003, C-J Hilder, www.AstronomyInYourHands.com

Widmo mocy $P(f)$

(widmowa gęstość energii)

Względny udział składowych harmoniczných w szeregu czasowym. Np.:

- widmo mocy sygnału sinusoidalnego będzie miało *pik* na częstotliwości tego sinusa; kombinacja liniowa dwóch różnych sinusów \rightarrow dwa *piki*.

Widmo mocy $P(f)$

(widmowa gęstość energii)

Względny udział składowych harmoniczných w szeregu czasowym. Np.:

- widmo mocy sygnału sinusoidalnego będzie miało *pik* na częstotliwości tego sinusa; kombinacja liniowa dwóch różnych sinusów \rightarrow dwa *piki*.
- Równy udział wszystkich częstotliwości \rightarrow płaskie widmo (biały szum).

Widmo mocy $P(f)$

(widmowa gęstość energii)

Względny udział składowych harmoniczných w szeregu czasowym. Np.:

- widmo mocy sygnału sinusoidalnego będzie miało *pik* na częstotliwości tego sinusa; kombinacja liniowa dwóch różnych sinusów \rightarrow dwa *piki*.
- Równy udział wszystkich częstotliwości \rightarrow płaskie widmo (biały szum).
- Widma postaci $P(f) \propto \frac{1}{f^\beta}$ nazywamy kolorowym szumem:
 - $\beta = -2$ — fioletowy szum
 - $\beta = -1$ — niebieski szum
 - $\beta = 0$ — biały szum
 - $\beta = 1$ — różowy szum
 - $\beta = 2$ — czerwony szum
 - $\beta > 2$ — czarny szum

Widmo mocy $P(f)$

(widmowa gęstość energii)

Względny udział składowych harmoniczných w szeregu czasowym. Np.:

- widmo mocy sygnału sinusoidalnego będzie miało *pik* na częstotliwości tego sinusa; kombinacja liniowa dwóch różnych sinusów \rightarrow dwa *piki*.
- Równy udział wszystkich częstotliwości \rightarrow płaskie widmo (biały szum).
- Widma postaci $P(f) \propto \frac{1}{f^\beta}$ nazywamy kolorowym szumem:
 - $\beta = -2$ — fioletowy szum
 - $\beta = -1$ — niebieski szum
 - $\beta = 0$ — biały szum
 - $\beta = 1$ — różowy szum
 - $\beta = 2$ — czerwony szum
 - $\beta > 2$ — czarny szum
- Inne formy szumu, np. szary szum.

Periodogram Lomb-Scargle'a (LSP)

Szereg czasowy $\{(t_k, x_k)\}_{k=1}^N$ taki, że $\delta t_k = t_k - t_{k-1} \neq \text{const.}$

Periodogram Lomb-Scargle'a (LSP)

Szereg czasowy $\{(t_k, x_k)\}_{k=1}^N$ taki, że $\delta t_k = t_k - t_{k-1} \neq \text{const.}$

Niech $\delta t_k = m \cdot \delta t = m \cdot \min\{\delta t_k\}_{k=2}^N$, $m \in \mathbb{N}$.

Periodogram Lomb-Scargle'a (LSP)

Szereg czasowy $\{(t_k, x_k)\}_{k=1}^N$ taki, że $\delta t_k = t_k - t_{k-1} \neq \text{const.}$

Niech $\delta t_k = m \cdot \delta t = m \cdot \min\{\delta t_k\}_{k=2}^N$, $m \in \mathbb{N}$.

$$P(f) = \frac{1}{2S^2} \left[\frac{\left(\sum_{k=1}^N (x_k - \bar{x}) \cos[2\pi f(t_k - \tau)] \right)^2}{\sum_{k=1}^N \cos^2[2\pi f(t_k - \tau)]} + \frac{\left(\sum_{k=1}^N (x_k - \bar{x}) \sin[2\pi f(t_k - \tau)] \right)^2}{\sum_{k=1}^N \sin^2[2\pi f(t_k - \tau)]} \right]$$

często wyrażony przez $\omega = 2\pi f$; \bar{x} , S^2 — średnia i wariancja próbki

$$\tau \equiv \tau(f) = \frac{1}{4\pi f} \arctan \left[\frac{\sum_{k=1}^N \sin(4\pi f t_k)}{\sum_{k=1}^N \cos(4\pi f t_k)} \right]$$

Periodogram Lomb-Scargle'a (LSP)

Szereg czasowy $\{(t_k, x_k)\}_{k=1}^N$ taki, że $\delta t_k = t_k - t_{k-1} \neq \text{const.}$

Niech $\delta t_k = m \cdot \delta t = m \cdot \min\{\delta t_k\}_{k=2}^N$, $m \in \mathbb{N}$.

$$P(f) = \frac{1}{2S^2} \left[\frac{\left(\sum_{k=1}^N (x_k - \bar{x}) \cos[2\pi f(t_k - \tau)] \right)^2}{\sum_{k=1}^N \cos^2[2\pi f(t_k - \tau)]} + \frac{\left(\sum_{k=1}^N (x_k - \bar{x}) \sin[2\pi f(t_k - \tau)] \right)^2}{\sum_{k=1}^N \sin^2[2\pi f(t_k - \tau)]} \right]$$

często wyrażony przez $\omega = 2\pi f$; \bar{x} , S^2 — średnia i wariancja próbki

$$\tau \equiv \tau(f) = \frac{1}{4\pi f} \arctan \left[\frac{\sum_{k=1}^N \sin(4\pi f t_k)}{\sum_{k=1}^N \cos(4\pi f t_k)} \right]$$

$$f_{\min} = \frac{1}{t_{\max} - t_{\min}}, \quad f_{\max} = \frac{1}{2\delta t};$$

Periodogram Lomb-Scargle'a (LSP)

Szereg czasowy $\{(t_k, x_k)\}_{k=1}^N$ taki, że $\delta t_k = t_k - t_{k-1} \neq \text{const.}$

Niech $\delta t_k = m \cdot \delta t = m \cdot \min\{\delta t_k\}_{k=2}^N$, $m \in \mathbb{N}$.

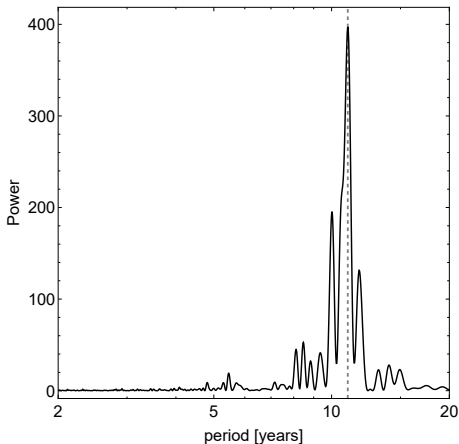
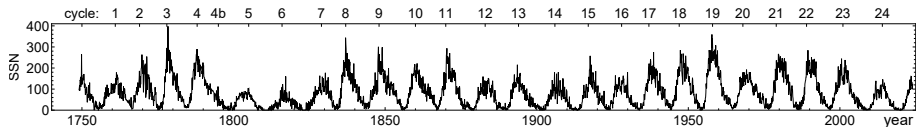
$$P(f) = \frac{1}{2S^2} \left[\frac{\left(\sum_{k=1}^N (x_k - \bar{x}) \cos[2\pi f(t_k - \tau)] \right)^2}{\sum_{k=1}^N \cos^2[2\pi f(t_k - \tau)]} + \frac{\left(\sum_{k=1}^N (x_k - \bar{x}) \sin[2\pi f(t_k - \tau)] \right)^2}{\sum_{k=1}^N \sin^2[2\pi f(t_k - \tau)]} \right]$$

często wyrażony przez $\omega = 2\pi f$; \bar{x} , S^2 — średnia i wariancja próbki

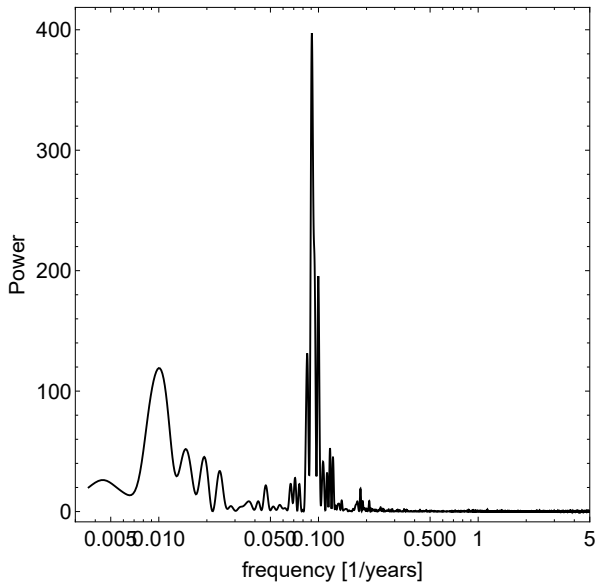
$$\tau \equiv \tau(f) = \frac{1}{4\pi f} \arctan \left[\frac{\sum_{k=1}^N \sin(4\pi f t_k)}{\sum_{k=1}^N \cos(4\pi f t_k)} \right]$$

$$f_{\min} = \frac{1}{t_{\max} - t_{\min}}, \quad f_{\max} = \frac{1}{2\delta t}; \quad N_{\text{periods}} = n_0 \frac{f_{\max}}{f_{\min}} \quad \text{gdzie } n_0 \approx 10$$

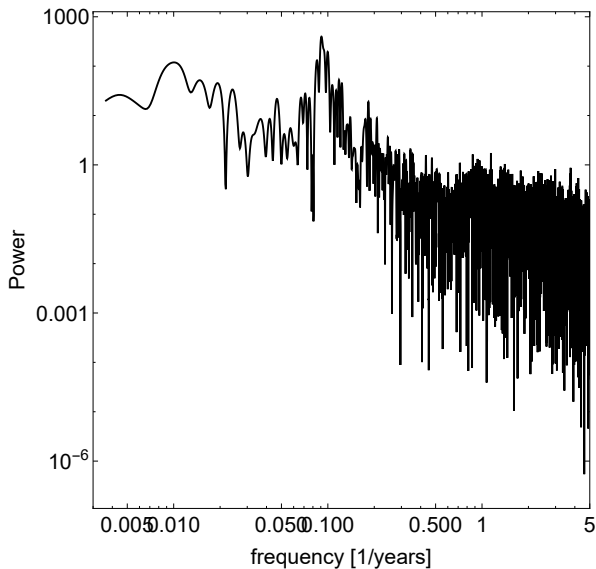
Cykl słoneczny (*sunspot number, SSN*)



Widmo mocy — $P(f)$



Widmo mocy — $P(f)$



Zadanie 14 (*analyczne*)

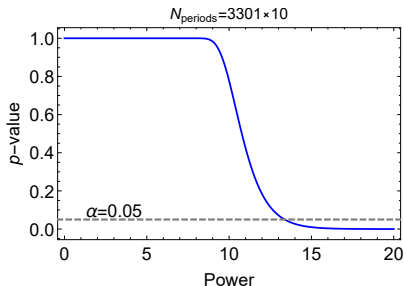
Założyć, że $\delta t_k = \text{const}$. Uprościć stosowne sumy w LSP. Pokazać, że dla jednorodnie próbkowanego szeregu czasowego LSP jest w *przybliżeniu* zgodny z periodogramem fourierowskim.

Istotność względem białego szumu (p -wartość)

$$p = 1 - (1 - e^{-x})^{N_{\text{periods}}}$$

gdzie x to wartość mocy $P(f)$ periodogramu na zadanej częstotliwości f .

Istotne statystycznie są piki, dla których $p < \alpha$. Na przykład:

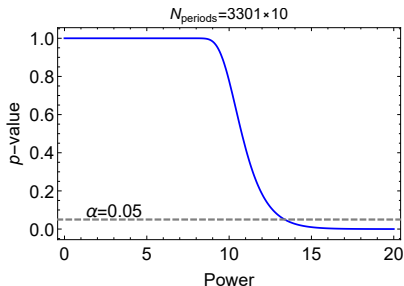


Istotność względem białego szumu (p -wartość)

$$p = 1 - (1 - e^{-x})^{N_{\text{periods}}}$$

gdzie x to wartość mocy $P(f)$ periodogramu na zadanej częstotliwości f .

Istotne statystycznie są piki, dla których $p < \alpha$. Na przykład:



UWAGA!

To jest istotność istnienia w sygnale pewnej okresowości (hipoteza alternatywna) wobec hipotezy zerowej, że sygnał jest białym szumem.

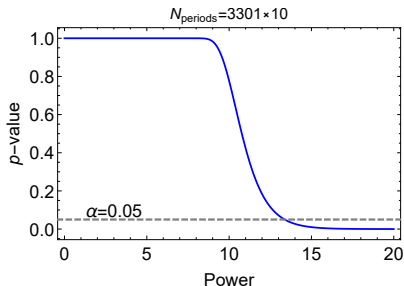
<https://docs.astropy.org/en/stable/timeseries/lombscargle.html>

Istotność względem białego szumu (p -wartość)

$$p = 1 - (1 - e^{-x})^{N_{\text{periods}}}$$

gdzie x to wartość mocy $P(f)$ periodogramu na zadanej częstotliwości f .

Istotne statystycznie są piki, dla których $p < \alpha$. Na przykład:



UWAGA!

To jest istotność istnienia w sygnale pewnej okresowości (hipoteza alternatywna) wobec hipotezy zerowej, że sygnał jest białym szumem.

<https://docs.astropy.org/en/stable/timeseries/lombscargle.html>

Jeśli hipotezą zerową jest np. $P(f) \propto 1/f^2$, to niemal cały periodogram będzie istotny statystycznie. Należy wtedy użyć metod Monte Carlo do wyznaczenia istotności.

Dopasowanie krzywych do danych

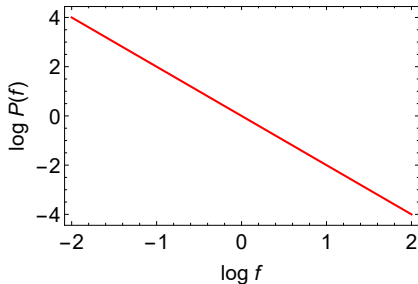
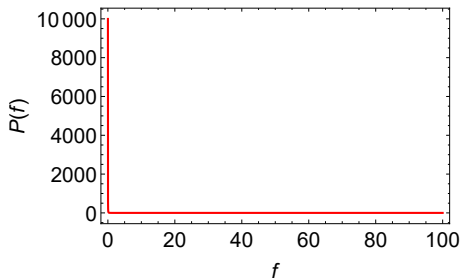
Prawo potęgowe (*power law*, PL):

$$P(f) = \frac{A}{f^\beta}$$

Dopasowanie krzywych do danych

Prawo potęgowe (*power law*, PL):

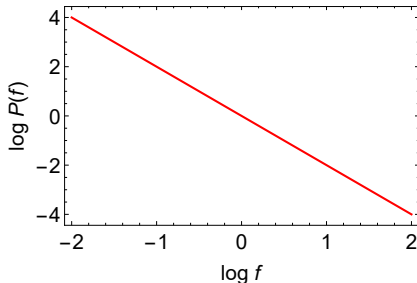
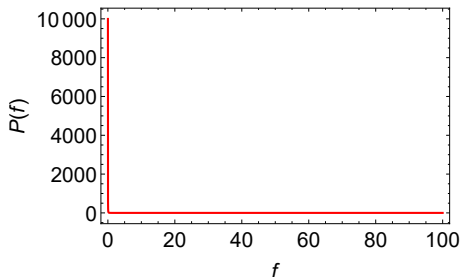
$$P(f) = \frac{A}{f^\beta}$$



Dopasowanie krzywych do danych

Prawo potęgowe (*power law*, PL):

$$P(f) = \frac{A}{f^\beta}$$

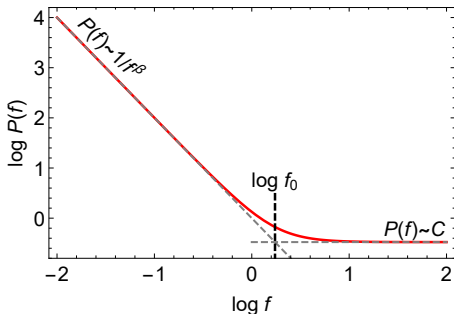


$$y \equiv \log P(f) = -\beta \log f + \log A \equiv -\beta x + \alpha$$

Dopasowanie krzywych do danych

Prawo potęgowe + poissonowski poziom szumu (PLC):

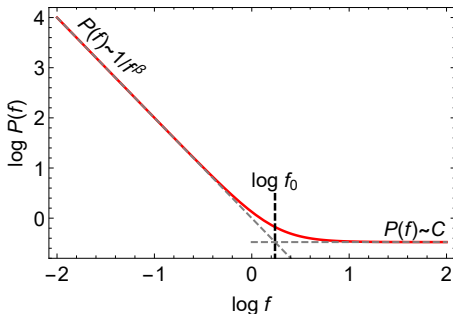
$$P(f) = \frac{A}{f^\beta} + C$$



Dopasowanie krzywych do danych

Prawo potęgowe + poissonowski poziom szumu (PLC):

$$P(f) = \frac{A}{f^\beta} + C$$



$$y \equiv \log P(f) = \log \left(\frac{A}{f^\beta} + C \right)$$

$$y \equiv \log P(f) = \log \left(\frac{A}{f^\beta} + C \right)$$

$$y \equiv \log P(f) = \log \left(\frac{A}{f^\beta} + C \right)$$

$$x = \log f$$

$$y \equiv \log P(f) = \log \left(\frac{A}{f^\beta} + C \right)$$

$$x = \log f$$

$$f = 10^x$$

$$y \equiv \log P(f) = \log \left(\frac{A}{f^\beta} + C \right)$$

$$x = \log f$$

$$f = 10^x$$

$$y(x) = \log \left(\frac{A}{10^{\beta x}} + C \right)$$

Dopasowanie krzywych do danych

$$y \equiv \log P(f) = \log \left(\frac{A}{f^\beta} + C \right)$$

$$x = \log f$$

$$f = 10^x$$

$$y(x) = \log \left(\frac{A}{10^{\beta x}} + C \right)$$

f_0 takie, że

$$\frac{A}{f_0^\beta} = C \Rightarrow$$

$$f_0 = \left(\frac{A}{C} \right)^{1/\beta}$$

Dopasowanie krzywych do danych

$$y \equiv \log P(f) = \log \left(\frac{A}{f^\beta} + C \right)$$

$$x = \log f$$

$$f = 10^x$$

$$y(x) = \log \left(\frac{A}{10^{\beta x}} + C \right)$$

f_0 takie, że

$$\frac{A}{f_0^\beta} = C \Rightarrow$$

$$f_0 = \left(\frac{A}{C} \right)^{1/\beta}$$

$$(x_i, y_i) = (\log f_i, \log P(f_i))$$

Poissonowski poziom szumu

Szum będący inherentnym wynikiem niepewności pomiarowych obecnych w samym szeregu czasowym:

$$P_{\text{Poisson}} = \frac{1}{2S^2} \frac{1}{N} \sum_{i=1}^N \Delta x_i^2$$

Poissonowski poziom szumu

Szum będący inherentnym wynikiem niepewności pomiarowych obecnych w samym szeregu czasowym:

$$P_{\text{Poisson}} = \frac{1}{2S^2} \frac{1}{N} \sum_{i=1}^N \Delta x_i^2$$

Ale — po zlogarytmowaniu (por. slajdy 19–21) pojawia się przesunięcie wynikające z faktu, że logarytmowanie podlega rozkładowi χ_2^2 [Vaughan, A&A 431, 391-403 (2005)].

Praktycznie rzecz ujmując: od $\log P_{\text{Poisson}}$ należy **odjąć** czynnik $\gamma / \ln(10) = 0.25068 \dots$, gdzie $\gamma = 0.57721 \dots$ jest stałą Eulera-Mascheroniego.

Poissonowski poziom szumu

Szum będący inherentnym wynikiem niepewności pomiarowych obecnych w samym szeregu czasowym:

$$P_{\text{Poisson}} = \frac{1}{2S^2} \frac{1}{N} \sum_{i=1}^N \Delta x_i^2$$

Ale — po zlogarytmowaniu (por. slajdy 19–21) pojawia się przesunięcie wynikające z faktu, że logarytmowanie podlega rozkładowi χ_2^2 [Vaughan, A&A 431, 391-403 (2005)].

Praktycznie rzecz ujmując: od $\log P_{\text{Poisson}}$ należy **odjąć** czynnik $\gamma/\ln(10) = 0.25068\dots$, gdzie $\gamma = 0.57721\dots$ jest stałą Eulera-Mascheroniego.

Równoważnie: $\log P_{\text{Poisson}}$ obliczyć jw., zaś do $\log P(f)$ **dodać** $\gamma/\ln(10)$.

Poissonowski poziom szumu

Szum będący inherentnym wynikiem niepewności pomiarowych obecnych w samym szeregu czasowym:

$$P_{\text{Poisson}} = \frac{1}{2S^2} \frac{1}{N} \sum_{i=1}^N \Delta x_i^2$$

Ale — po zlogarytmowaniu (por. slajdy 19–21) pojawia się przesunięcie wynikające z faktu, że logarytmowanie podlega rozkładowi χ_2^2 [Vaughan, A&A 431, 391-403 (2005)].

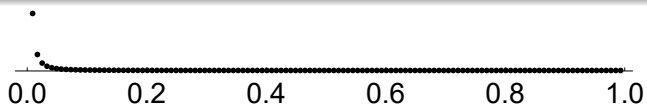
Praktycznie rzecz ujmując: od $\log P_{\text{Poisson}}$ należy **odjąć** czynnik $\gamma / \ln(10) = 0.25068 \dots$, gdzie $\gamma = 0.57721 \dots$ jest stałą Eulera-Mascheroniego.

Równoważnie: $\log P_{\text{Poisson}}$ obliczyć jw., zaś do $\log P(f)$ **dodać** $\gamma / \ln(10)$.

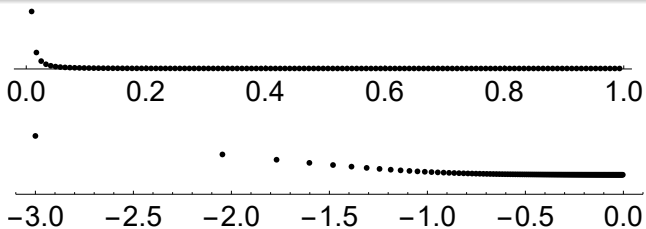
Powyższy wzór na P_{Poisson} jest zgodny z normalizacją LSP ze slajdu 13.

Implementacja w astropy jest inna; przeskalowanie jest empiryczne.

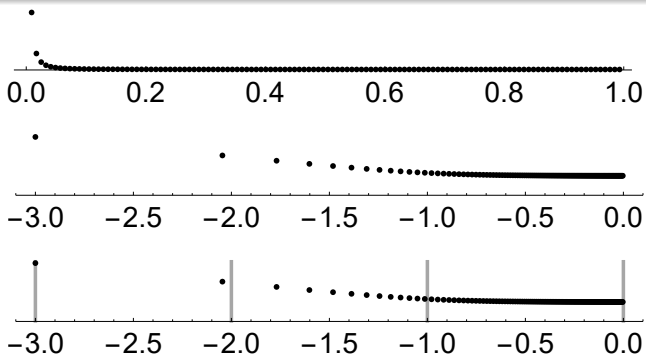
Kształt widma mocy z LSP



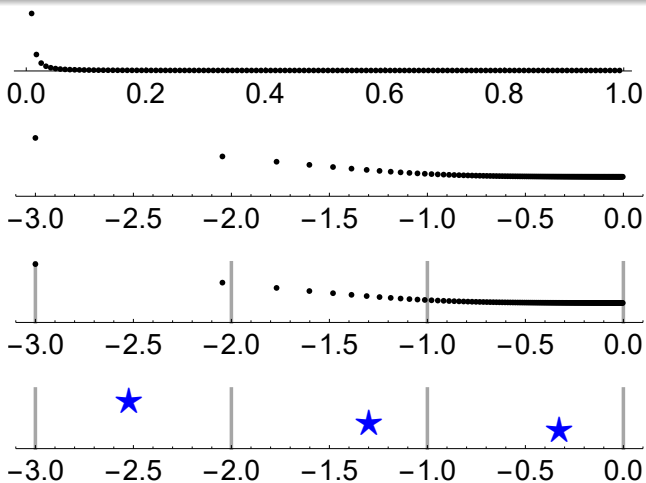
Kształt widma mocy z LSP



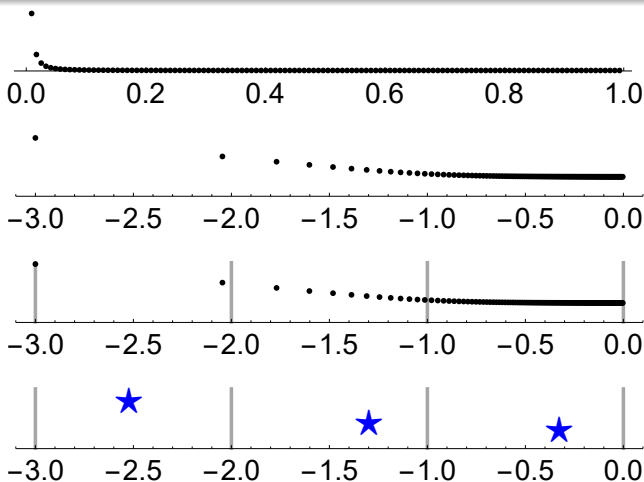
Kształt widma mocy z LSP



Kształt widma mocy z LSP



Kształt widma mocy z LSP



Reprezentatywny $\log f$ w binie — jako średnia geom.: $\sqrt[n]{x_1 x_2 \dots x_n}$
Reprezentatywny $\log P(f)$ w binie — jako średnia arytmetyczna.

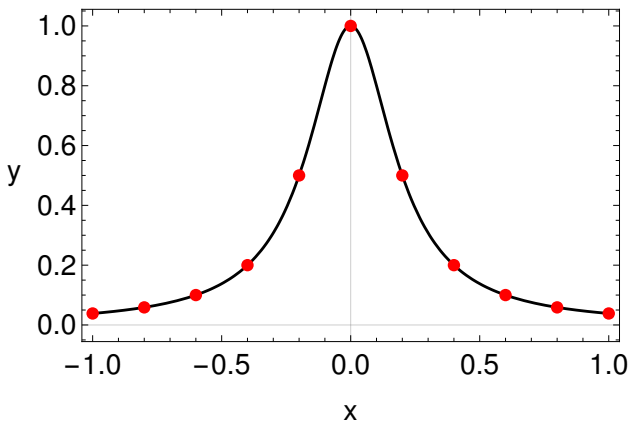
Jak wybrać lepszy model?

Mamy zestaw danych pomiarowych/obserwacyjnych. Mamy kilka modeli teoretycznych/fenomenologicznych, które mogą te dane opisywać.

Pytanie:

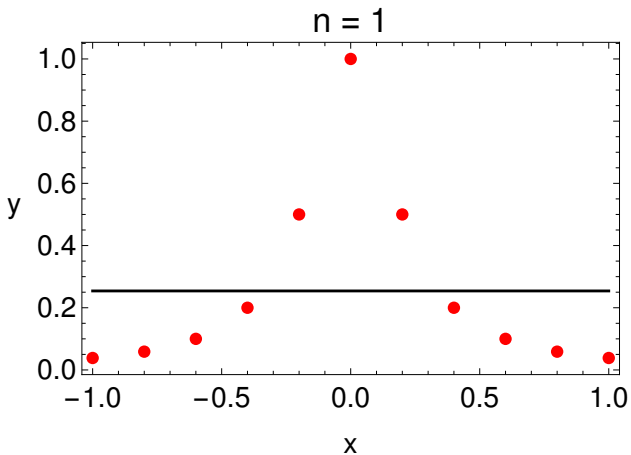
Który model jest najlepszy **oraz najprostszy**?

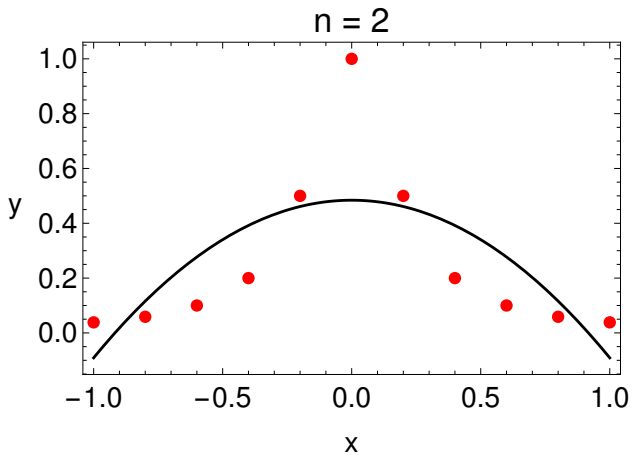
$$y(x) = \frac{1}{1 + 25x^2}$$

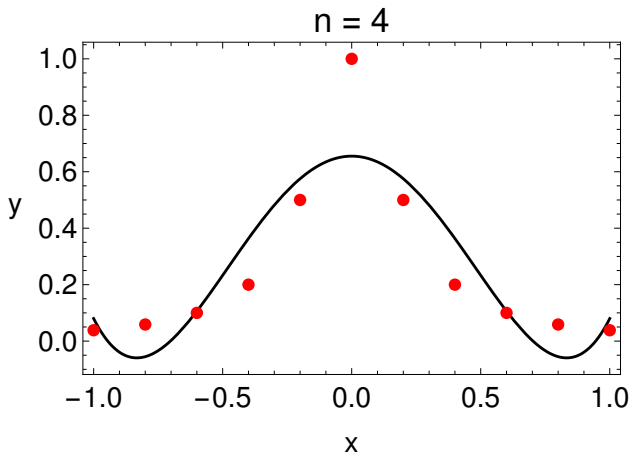


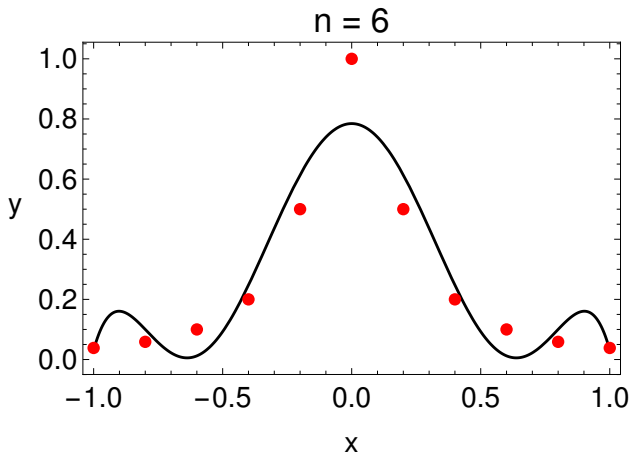
$$f_n(x) = a_0 + a_1x + \dots + a_nx^n$$

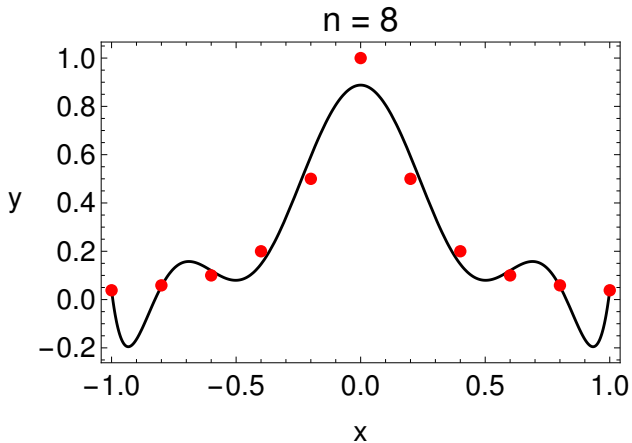
$$f_n(x) = a_0 + a_1x + \dots + a_nx^n$$





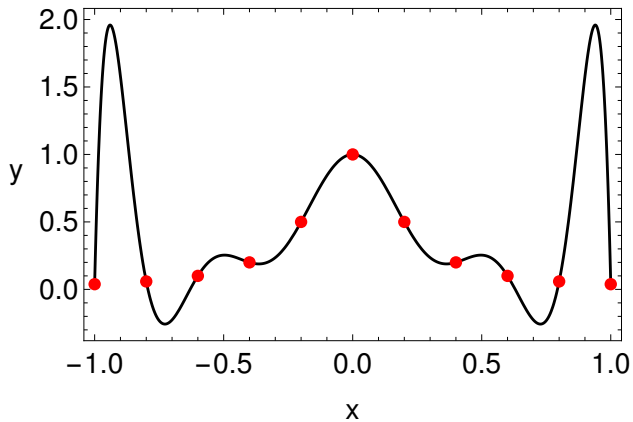






Oscylacje Runge'go:

$n = 10$



Kryteria informacyjne: AIC oraz BIC

N punktów pomiarowych (x_i, y_i) ; model z p parametrami.

- Akaike Information Criterion (Akaike 1974):

$$AIC = 2p - 2\mathcal{L}_{\max}$$

- Bayesian Information Criterion (Schwarz 1978):

$$BIC = p \ln N - 2\mathcal{L}_{\max}$$

Kryteria informacyjne: AIC oraz BIC

N punktów pomiarowych (x_i, y_i) ; model z p parametrami.

- Akaike Information Criterion (Akaike 1974):

$$AIC = 2p - 2\mathcal{L}_{\max}$$

- Bayesian Information Criterion (Schwarz 1978):

$$BIC = p \ln N - 2\mathcal{L}_{\max}$$

Caveat emptor

Niektóre oprogramowania zamiast maksymalizować \mathcal{L} , minimalizują oraz zwracają $-\mathcal{L}$ (np. `mclust`).

Kryteria informacyjne: AIC oraz BIC

Sprawdzamy pewną liczbę modeli. Niech jeden z nich ma AIC_{\min} (albo BIC_{\min}). O ile jest lepszy niż pozostałe?

Kryteria informacyjne: AIC oraz BIC

Sprawdzamy pewną liczbę modeli. Niech jeden z nich ma AIC_{\min} (albo BIC_{\min}). O ile jest lepszy niż pozostałe?

$$\Delta AIC_j = AIC_j - AIC_{\min} \quad \text{oraz} \quad \Delta BIC_j = BIC_j - BIC_{\min}$$

Kryteria informacyjne: AIC oraz BIC

Sprawdzamy pewną liczbę modeli. Niech jeden z nich ma AIC_{\min} (albo BIC_{\min}). O ile jest lepszy niż pozostałe?

$$\Delta AIC_j = AIC_j - AIC_{\min} \quad \text{oraz} \quad \Delta BIC_j = BIC_j - BIC_{\min}$$

AIC

- $0 < \Delta_j < 2$ — substantial support (for the j -th model)
- $2 < \Delta_j < 4$ — strong support
- $4 < \Delta_j < 7$ — considerably less support
- $7 < \Delta_j < 10$ — almost no support
- $\Delta_j > 10$ — essentially no support

BIC

- $0 < \Delta_j < 2$ — substantial support
- $2 < \Delta_j < 6$ — evidence against the j -th model
- $6 < \Delta_j < 10$ — evidence is strong
- $\Delta_j > 10$ — very strong evidence

Kryteria informacyjne: AIC oraz BIC

Sprawdzamy pewną liczbę modeli. Niech jeden z nich ma AIC_{\min} (albo BIC_{\min}). O ile jest lepszy niż pozostałe?

$$\Delta AIC_j = AIC_j - AIC_{\min} \quad \text{oraz} \quad \Delta BIC_j = BIC_j - BIC_{\min}$$

AIC

- $0 < \Delta_j < 2$ — substantial support (for the j -th model)
- $2 < \Delta_j < 4$ — strong support
- $4 < \Delta_j < 7$ — considerably less support
- $7 < \Delta_j < 10$ — almost no support
- $\Delta_j > 10$ — essentially no support

BIC

- $0 < \Delta_j < 2$ — substantial support
- $2 < \Delta_j < 6$ — evidence against the j -th model
- $6 < \Delta_j < 10$ — evidence is strong
- $\Delta_j > 10$ — very strong evidence

Caveat emptor

Różne oprogramowania używają różnych normalizacji — wartości AIC czy BIC z jednego często nie można przyrównać do wartości z innego!

Jeśli $N/p < 40$, to zaleca się użycie AIC z poprawką, tj.

$$AIC_c = 2p - 2\mathcal{L}_{\max} + \frac{2(p+1)(p+2)}{N-p-2} \equiv AIC + \frac{2(p+1)(p+2)}{N-p-2}$$

Jeśli N/p jest duże, to $AIC_c \approx AIC$.

Jeśli $N/p < 40$, to zaleca się użycie AIC z poprawką, tj.

$$AIC_c = 2p - 2\mathcal{L}_{\max} + \frac{2(p+1)(p+2)}{N-p-2} \equiv AIC + \frac{2(p+1)(p+2)}{N-p-2}$$

Jeśli N/p jest duże, to $AIC_c \approx AIC$.

W zasadzie nic nie stoi na przeszkodzie, by zawsze używać AIC_c .

Jeśli $N/p < 40$, to zaleca się użycie AIC z poprawką, tj.

$$AIC_c = 2p - 2\mathcal{L}_{\max} + \frac{2(p+1)(p+2)}{N-p-2} \equiv AIC + \frac{2(p+1)(p+2)}{N-p-2}$$

Jeśli N/p jest duże, to $AIC_c \approx AIC$.

W zasadzie nic nie stoi na przeszkodzie, by zawsze używać AIC_c .

Przedziały interpretacyjne dla $\Delta AIC_{c,j}$ są takie same jak dla ΔAIC_j .

Kontekst astrofizyczny

Krzywe zmian blasku blazarów w zakresie promieniowania gamma pochodzą głównie z relatywistycznego dżetu. Jeśli założymy, że zmienność dżetu jest powiązana ze zmiennością dysku akrecyjnego, który z kolei jest opisany modelem Shakury-Sunyaeva [[A&A 24, 337-355 \(1973\)](#)], tj. jest to płyn o lepkości α , to fizyka akrecji będzie determinowała obserwowaną zmienność.

Kontekst astrofizyczny

Krzywe zmian blasku blazarów w zakresie promieniowania gamma pochodzą głównie z relatywistycznego dżetu. Jeśli założymy, że zmienność dżetu jest powiązana ze zmiennością dysku akrecyjnego, który z kolei jest opisany modelem Shakury-Sunyaeva [[A&A 24, 337-355 \(1973\)](#)], tj. jest to płyn o lepkości α , to fizyka akrecji będzie determinowała obserwowaną zmienność. Załóżmy, że α fluktuuje w czasie t i w zależności od odległości radialnej r od czarnej dziury [[Lyubarskii, MNRAS 292, 679-685 \(1997\)](#)]:

$$\alpha = \alpha_0[1 + \xi(t, r)]$$

gdzie $\xi \ll 1$ oraz

$$\sqrt{\langle \xi^2 \rangle} \propto r^b$$

Kontekst astrofizyczny

Krzywe zmian blasku blazarów w zakresie promieniowania gamma pochodzą głównie z relatywistycznego dżetu. Jeśli założymy, że zmienność dżetu jest powiązana ze zmiennością dysku akrecyjnego, który z kolei jest opisany modelem Shakury-Sunyaeva [A&A 24, 337-355 (1973)], tj. jest to płyn o lepkości α , to fizyka akrecji będzie determinowała obserwowaną zmienność. Załóżmy, że α fluktuuje w czasie t i w zależności od odległości radialnej r od czarnej dziury [Lyubarskii, MNRAS 292, 679-685 (1997)]:

$$\alpha = \alpha_0[1 + \xi(t, r)]$$

gdzie $\xi \ll 1$ oraz

$$\sqrt{\langle \xi^2 \rangle} \propto r^b$$

Wtedy

$$P(f) \propto f^{-(1+4b/3)}$$

co w połączeniu z β wyznaczonym z widma mocy (LSP) daje informacje nt. procesów rządzących zmiennością blazarów.

Projekt na zaliczenie (sprawozdanie)

Zadanie 15

Zbadać kształt widma mocy [przypisanego blazara](#).

- 1 Skonstruować LSP wg implementacji pakietu `astropy`. Wyznaczyć okres najsilniejszej składowej okresowej oraz określić jej istotność statystyczną. Wziąć pod uwagę niepewności obserwacji w krzywych zmian blasku.
- 2 Zbinować LSP w przestrzeni logarytmicznej ($\log f, \log P(f)$).
- 3 Dopasować do zbinowanego LSP modele PL i PLC. W przypadku PLC wyznaczyć skalę czasu $T = 1/f_0$ oraz jej niepewność. Skorzystać z AIC_c oraz BIC do wybrania odp. modelu. Czy wskazania są zgodne?

Zbadać zarówno krzywe zmian blasku, jak i zlogarytmowane (prawo propagacji błędów!) krzywe zmian blasku (por. slajd 7).

Zaznaczyć na wykresach: surowe i zbinowane LSP, dopasowania PL i PLC, f_0 wraz z niepewnością.

Korzystając z wartości β wyznaczyć b (zob. slajd 35) wraz z niepewnością.

Wykres referencyjny

