

Stacjonarne procesy gaussowskie, czyli o związkach pomiędzy zwykłą i częściową funkcją autokorelacji

Łukasz Dębowski

ldebowsk@ipipan.waw.pl



Instytut Podstaw Informatyki PAN

Zarys referatu

- Co to są procesy gaussowskie?

rozkład normalny, macierz kowariancji,
najlepsze liniowe predyktory,
warunek istnienia

- Procesy stacjonarne:

funkcja autokorelacji, funkcja częściowej autokorelacji,
algorytm Durбина-Levinsona,
interpretacja teorioinformacyjna

- Wzory na sumę autokorelacji.

Miara prawdopodobieństwa

$\mathcal{J} \subset 2^\Omega$ jest **σ -ciałem**, gdy

1. $\Omega \in \mathcal{J}$,
2. $A \in \mathcal{J} \implies \Omega \setminus A \in \mathcal{J}$,
3. $(\forall n \in \mathbb{N} : A_n \in \mathcal{J}) \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{J}$.

$P : \mathcal{J} \rightarrow [0, 1]$ jest **miarą prawdopodobieństwa**, gdy

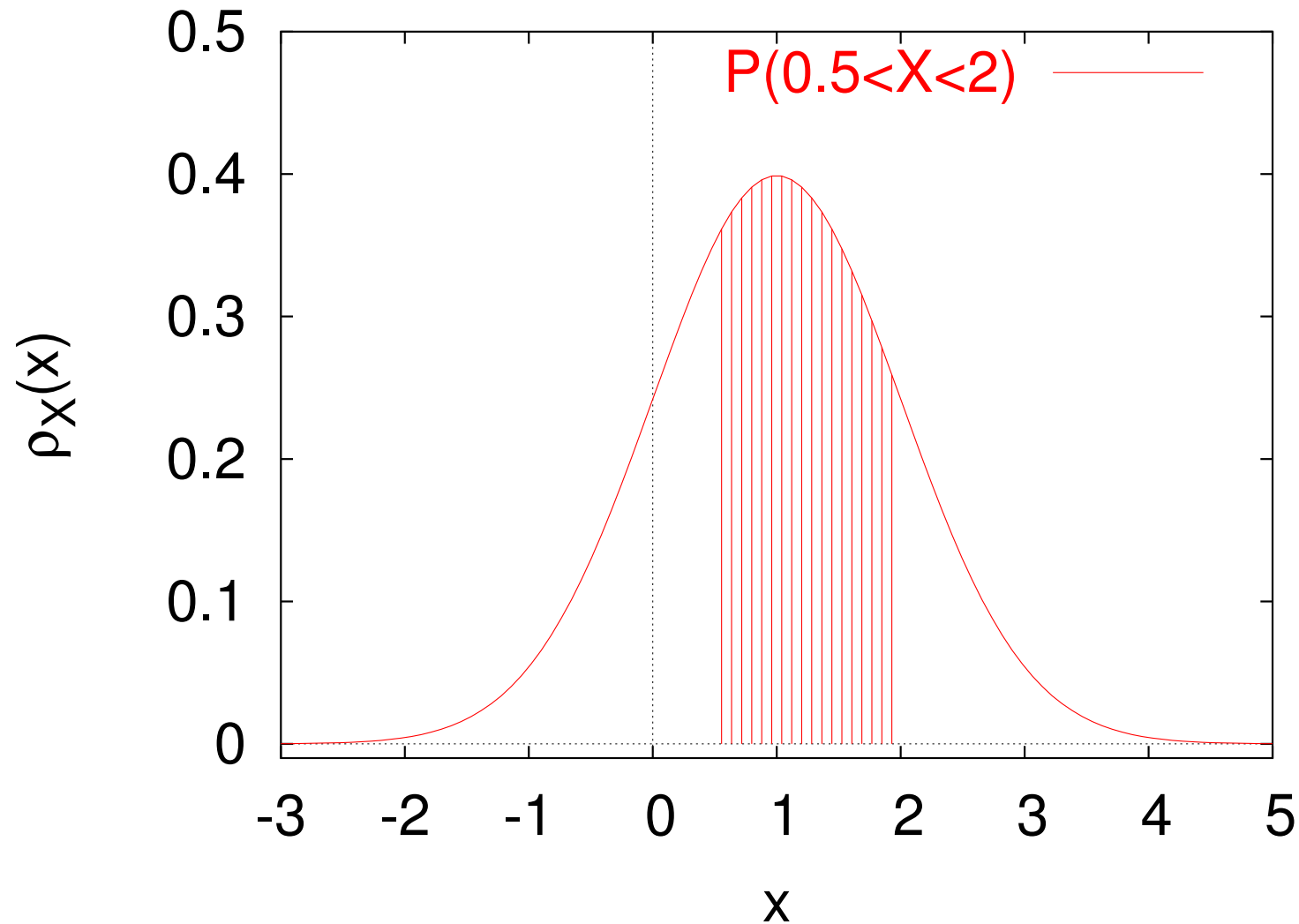
1. $P(\Omega) = 1$,
2. $P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$ dla rozłącznych A_n .

$X : \Omega \rightarrow \mathbb{B}$ jest **zmienną losową**. Definiujemy zdarzenie

$$(X \in \mathbb{A}) := X^{-1}(\mathbb{A}).$$

$P(X \in \mathbb{A})$ jest określone wtw., gdy $X^{-1}(\mathbb{A}) \in \mathcal{J}$.

Gęstość prawdopodobieństwa (rozkład zmiennej)



Rozkłady zmiennych

- miara prawdopodobieństwa:

$$P(X \in \mathbb{A}) := \int_{\mathbb{A}} \rho_X(x) dx = P(X^{-1}(\mathbb{A})) = \int_{X^{-1}(\mathbb{A})} dP$$

- wartość oczekiwana:

$$\langle f(X) \rangle := \int f(x) \rho_X(x) dx = \int f(X) dP$$

$P(A)$ może być określone także, gdy $A \neq (X \in \mathbb{A})$.

Definicje z dP są wygodniejsze dla dużych zbiorów zmiennych.

$$\int F dP := \sup_{\text{rozbicie}} \sum_{A \in \text{rozbicie}} P(A) \inf_{\text{argument} \in A} F(\text{argument})$$

Procesy stochastyczne

Proces to dowolny zbiór zmiennych, np. **nieprzeliczalny**, określonych na tej samej przestrzeni zdarzeń Ω .

Gęstości tych zmiennych spełniają warunki zgodności:

$$\rho_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int \rho_{X_1, \dots, X_n, X_{n+1}}(x_1, \dots, x_n, x_{n+1}) dx_{n+1}.$$

tw.tw. Kołmogorowa o procesie i rozszerzeniu:

Kandydaci na gęstości $\rho_{X_1, X_2, \dots, X_n}$, $n \in \mathbb{N}$, spełniający warunki zgodności definiują miarę prawdopodobieństwa P na pewnym σ -ciele.

Procesy gaussowskie

Wielowymiarowy rozkład normalny (Gaussa) to najprostszy algebraicznie model zależnych zmiennych losowych:

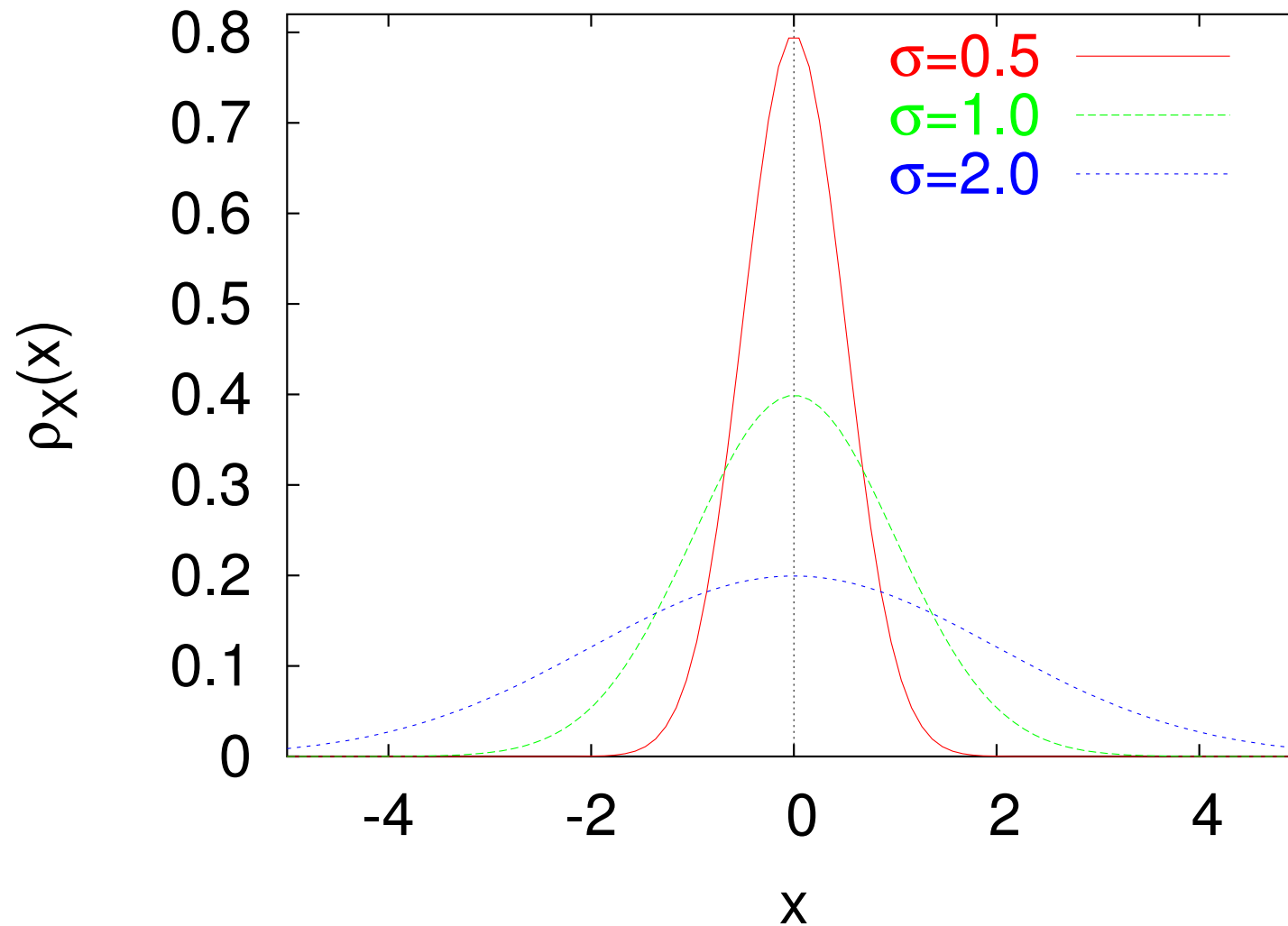
- Łatwo podać wielowymiarowe gęstości.
- Łatwo **wyrugować** wielowymiarowe gęstości z rozważań.

Do tego:

- Dowolny ciąg **kombinacji liniowych** zmiennych o wielowymiarowym rozkładzie normalnym ma wielowymiarowy rozkład normalny.
- Dowolny ciąg zmiennych o wielowymiarowym rozkładzie normalnym można wyrazić jako ciąg kombinacji liniowych **niezależnych** zmiennych losowych o rozkładzie normalnym.

Dowolny ciąg zmiennych gaussowskich można uzupełnić do **nieprzeliczalnego** procesu tworzącego **przestrzeń Hilberta**.

Jednowymiarowy rozkład Gaussa (normalny)



Jednowymiarowy rozkład Gaussa (normalny)

- gęstość prawdopodobieństwa:

$$\rho_X(x) = \frac{1}{[2\pi\sigma^2]^{d/2}} \exp\left[-\frac{1}{2d} \cdot \frac{|x|^2}{\sigma^2}\right]$$

- wariancja:

$$\sigma := \sqrt{\text{Var}(X)} \geq 0$$

$$\text{Var}(X) := \langle |X - \langle X \rangle|^2 \rangle = \langle |X|^2 \rangle - |\langle X \rangle|^2$$

Wielowymiarowy rozkład Gaussa (normalny)

- gęstość prawdopodobieństwa:

$$\rho_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{\exp \left[-\frac{1}{2d} \sum_{i=1}^n \sum_{j=1}^n x_i^* [\Gamma(n)^{-1}]_{ij} x_j \right]}{[(2\pi)^n \det \Gamma(n)]^{d/2}}$$

- macierz kowariancji (hermitowska, $\Gamma_{ij} = \Gamma_{ji}^*$):

$$\Gamma(n)_{ij} = \Gamma_{ij} = \text{Cov}(X_i; X_j)$$

$$\begin{aligned} \text{Cov}(X; Y) &:= \langle (X - \langle X \rangle)^* (Y - \langle Y \rangle) \rangle \\ &= \langle X^* Y \rangle - \langle X^* \rangle \langle Y \rangle \end{aligned}$$

$$\text{Cov}(X; X) = \text{Var}(X)$$

Analogia do $\sigma \geq 0$

Nie każda macierz hermitowska jest macierzą kowariancji.

Mamy

$$\text{Var} \left(\sum_{i=1}^n a_i^* X_i \right) = \sum_{i=1}^n \sum_{j=1}^n a_i^* \Gamma_{ij} a_j.$$

Macierz $\Gamma(n)$ musi być nieujemna określona, tzn.

$$\forall a_i \in \mathbb{C} : \sum_{i=1}^n \sum_{j=1}^n a_i^* \Gamma_{ij} a_j \geq 0.$$

Jest to warunek **konieczny i dostateczny** istnienia procesu.

Tylko jak go efektywnie sprawdzić?

Proces stacjonarny

Weźmy szereg czasowy czyli proces

$$X_{\mathbb{Z}} = \{X_i : i = \dots, -2, -1, 0, 1, 2, \dots\}.$$

Dla ułatwienia ograniczymy się do procesów stacjonarnych.

- proces stacjonarny:

$$\Gamma(n)_{ij} = \Gamma_{ij} = \text{Cov}(X_i; X_j) = \gamma(i - j)$$

- funkcja autokowariancji:

$$\gamma(n) := \text{Cov}(X_n; X_0)$$

Funkcja autokorelacji

- korelacja:

$$\text{Corr}(X, Y) := \frac{\text{Cov}(X; Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}, \quad |\text{Corr}(X, Y)| \leq 1$$

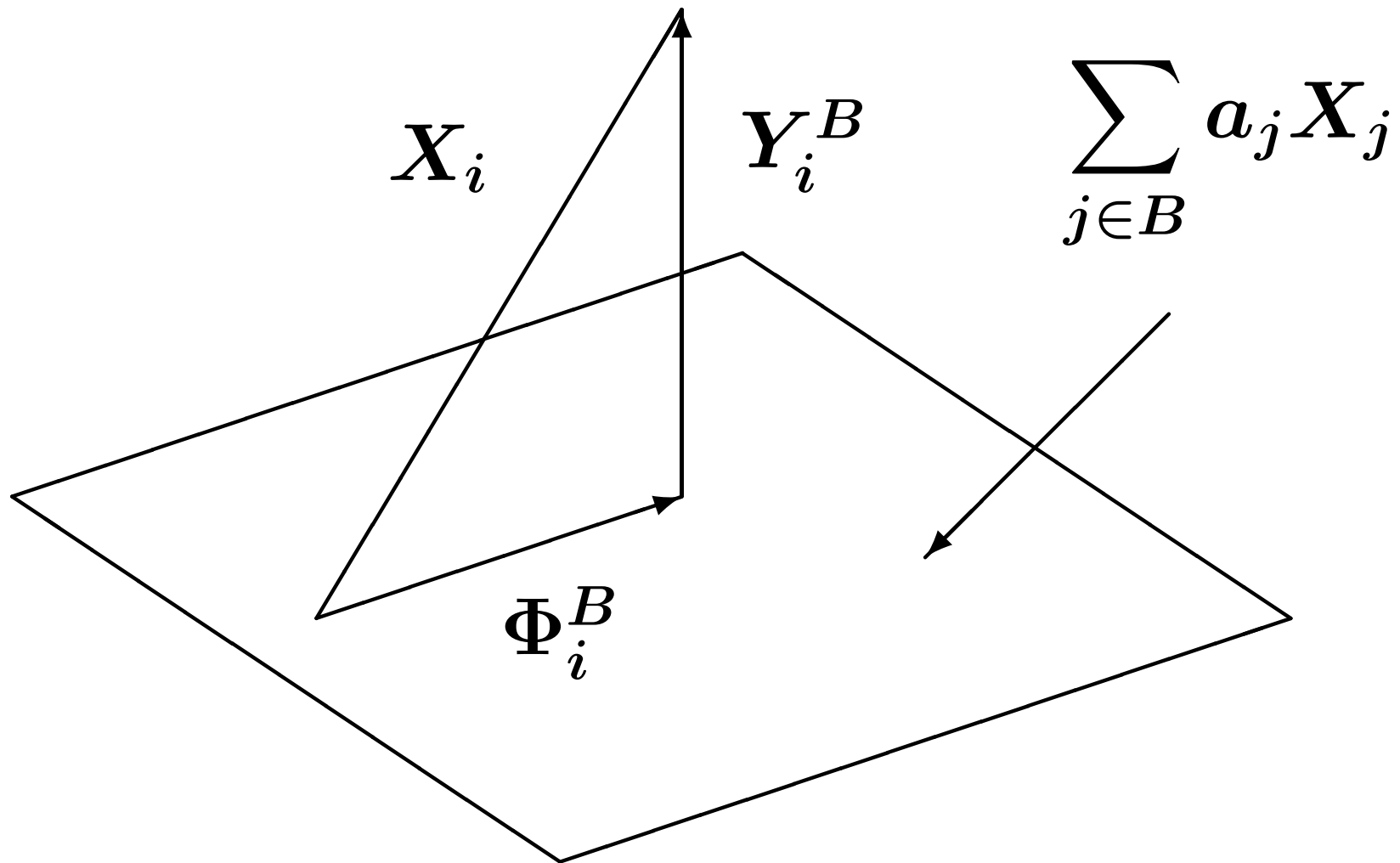
- funkcja autokorelacji (ACF):

$$\rho(n) := \text{Corr}(X_n; X_0) = \gamma(n)/\gamma(0)$$

- autokowariancja ma prosty estymator:

$$\gamma(n) \approx \frac{1}{N} \sum_{k=1}^{N-n} \left[x_{k+n} - \frac{1}{N} \sum_{j=1}^N x_j \right]^* \left[x_k - \frac{1}{N} \sum_{j=1}^N x_j \right]$$

Twierdzenie o rzucie ortogonalnym



Twierdzenie o rzucie ortogonalnym

Dla każdych zmiennych $(X_j)_{j \in B}$ istnieje jednoznaczny rozkład zmiennej X_i na **najlepszy predyktor** Φ_i^B i **innowację** Y_i^B :

$$X_i = \Phi_i^B + Y_i^B,$$

takie, że

$$\Phi_i^B = \sum_{j \in B} \phi_{ij}^B X_j, \quad \text{Cov}(Y_i^B; X_j) = 0, \quad j \in B,$$

$$\text{Var}(Y_i^B) = \min_{(a_j)_{j \in B}} \text{Var} \left(X_i - \sum_{j \in B} a_j X_j \right).$$

Dla danej macierzy kowariancji ϕ_{ij}^B oblicza się z algorytmu ortogonalizacji Grama-Schmidta.

Probabilistyczna niezależność

Zbiory zmiennych gaussowskich $(X_i)_{i \in A}$ oraz $(X_j)_{j \in B}$ są probabilistycznie niezależne wtedy i tylko wtedy, gdy

$$\text{Cov}(X_i; X_j) = 0, \quad i \in A, j \in B.$$

Wniosek:

Dla zmiennych gaussowskich innowacja Y_i^B jest probabilistycznie niezależna od $(X_j)_{j \in B}$.

„Niezwyczajna” funkcja autokorelacji

- (zwykła) funkcja autokorelacji (ACF):

$$\rho(n) := \text{Corr}(X_n; X_0) = \gamma(n)/\gamma(0)$$

- funkcja częściowej autokorelacji (PACF):

$$\begin{aligned}\alpha(n) &:= \text{Corr}(Y_n^{1:n-1}; Y_0^{1:n-1}) \\ &= \text{Corr}(X_n - \Phi_n^{1:n-1}; X_0 - \Phi_0^{1:n-1})\end{aligned}$$

$$n : m := \{n, n + 1, \dots, m\}$$

Warunek istnienia procesu — prościej

Istnieje **bijekcja** $(\rho(1), \dots, \rho(n)) \longleftrightarrow (\alpha(1), \dots, \alpha(n))$.

- Warunek dla ACF:

$$\forall a_i \in \mathbb{C} : \sum_{i=1}^n \sum_{j=1}^n a_i^* \rho(i-j) a_j \geq 0.$$

- Warunek dla PACF (Ramsey 1974):

$$|\alpha(n)| \leq 1 \quad \text{dla wszystkich } n \geq 1,$$
$$|\alpha(k)| = 1 \implies \alpha(n) = 0 \quad \text{dla } n > k.$$

To jest warunek dla PACF obliczonego z ACF.

To PACF a nie ACF jest „prostszy” obiektem.

Predyktory dla stacjonarnych

Równoważne definicji PACF:

$$\text{Cov}(Y_{n+1}^{2:n}; Y_1^{2:n}) = \alpha(n) \text{Cov}(Y_1^{2:n}; Y_1^{2:n}),$$

$$\text{Cov}(Y_{n+1}^{2:n}; X_1) = \alpha(n) \text{Cov}(Y_1^{2:n}; Y_1^{2:n}),$$

$$\text{Cov}(Y_{n+1}^{2:n}; X_1) = \alpha(n) \text{Cov}(Y_1^{2:n}; X_1).$$

Z ostatniego wynika:

$$\text{Cov}(Y_{n+1}^{2:n} - \alpha(n)Y_1^{2:n}; X_j) = 0, \quad j \in \{1, \dots, n\},$$

$$Y_{n+1}^{2:n} - \alpha(n)Y_1^{2:n} = \text{const} \cdot Y_{n+1}^{1:n}, \quad / - X_{n+1}$$

$$-\Phi_{n+1}^{2:n} - \alpha(n)Y_1^{2:n} = -\Phi_{n+1}^{1:n} + (\text{const} - 1) \cdot Y_{n+1}^{1:n},$$

gdzie $\text{const} = 1$, bo $Y_{n+1}^{1:n}$ liniowo niezależne od reszty.

Predyktory dla stacjonarnych

$$\Phi_{n+1}^{1:n} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$$

$$\Phi_0^{1:n} = \sum_{j=1}^n \phi_{nj}^* X_j$$

$$\phi_{nj} := \begin{cases} -1 & \text{dla } j = 0 \\ \phi_{n+1, n+1-j}^{1:n} = (\phi_{0j}^{1:n})^* & \text{dla } j \in \{1, \dots, n\} \\ 0 & \text{dla innych } j \end{cases}$$

$$v_n := \frac{\text{Var}(Y_n^{1:n-1})}{\text{Var}(X_n)}$$

Algorytm Durbina-Levinsona

$$v_0 = 1, \quad \alpha(1) = \rho(1)$$

$$v_n = [1 - |\alpha(n)|^2] v_{n-1}$$

$$\phi_{nj} = \phi_{n-1,j} - \alpha^*(n) \phi_{n-1,n-j}^*$$

Istnieje **bijekcja** $(\rho(1), \dots, \rho(n)) \longleftrightarrow (\alpha(1), \dots, \alpha(n))$:

$$\alpha(n+1) = \left[\rho(n+1) - \sum_{j=1}^n \phi_{nj}^* \rho(n+1-j) \right] / v_n$$

$$\rho(n+1) = v_n \alpha(n+1) + \sum_{j=1}^n \phi_{nj}^* \rho(n+1-j)$$

Interpretacja teorioinformacyjna

Entropia blokowa i informacje wzajemne:

$$\begin{aligned} H_X(n) &:= H(X_1, \dots, X_n) = nH_X(1) + \sum_{k=2}^n (n - k + 1)\Delta^2 H_X(k), \\ &= \frac{1}{2} [1 + d \log(2\pi)] n + \frac{d}{2} \log \det \Gamma(n) \\ I(X_1; X_n) &= -\frac{d}{2} \log [1 - |\rho(n-1)|^2], \\ I(X_1; X_n | X_{2:n-1}) &= -\Delta^2 H_X(n) = -\frac{d}{2} \log [1 - |\alpha(n-1)|^2], \\ \det \Gamma(n) &= \gamma(0)^n \prod_{k=1}^{n-1} [1 - |\alpha(k)|^2]^{n-k}. \end{aligned}$$

Częściowe podsumowanie

Stacjonarne procesy gaussowskie są prostym modelem zależnych zmiennych losowych.

Mimo że najłatwiejsza do estymacji jest **funkcja autokorelacji**, klasę stacjonarnych procesów gaussowskich najprościej i najogólniej parametryzuje **częściowa funkcja autokorelacji**.

W modelu tym istnieją proste algorytmy na obliczanie autokorelacji, najlepszych predyktorów i miar informacji.

Klasyczny problem:

Czy można zgrubnie scharakteryzować ogólny przebieg ACF dla danego PACF bez długotrwałego liczenia?

(Trochę na ten temat napisano, ale nie wyczerpano.)

Jeden z głównych moich rezultatów

Jeżeli $\sum_{k=1}^{\infty} |\alpha(k)| < \infty$, $|\alpha(k)| < 1$, to zachodzi

$$\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty, \quad (1)$$

$$\sum_{k=-n}^n |\rho(k)|^2 \leq \prod_{k=1}^n \left(\frac{1 + |\alpha(k)|}{1 - |\alpha(k)|} \right)^2, \quad (2)$$

$$\sum_{k=-\infty}^{\infty} \rho(k) e^{ik\omega} \in \left[\prod_{k=1}^{\infty} \frac{1 - |\alpha(k)|}{1 + |\alpha(k)|}, \prod_{k=1}^{\infty} \frac{1 + |\alpha(k)|}{1 - |\alpha(k)|} \right] \quad (3)$$

zaś, jeżeli $\alpha(k) \in \mathbb{R}$ dla wszystkich k , to

$$\sum_{k=-\infty}^{\infty} (\pm 1)^k \rho(k) = \prod_{k=1}^{\infty} \frac{1 + (\pm 1)^k \alpha(k)}{1 - (\pm 1)^k \alpha(k)}. \quad (4)$$

Kilka wniosków z (4)

Jeżeli $(\pm 1)^k \alpha(k) < 0$:

Zwiększenie $|\alpha(k)|$ powoduje zmniejszenie $\sum_{k=-\infty}^{\infty} (\pm 1)^k \rho(k)$.

Jeżeli $(\pm 1)^k \alpha(k) > 0$:

Zwiększenie $|\alpha(k)|$ powoduje zwiększenie $\sum_{k=-\infty}^{\infty} (\pm 1)^k \rho(k)$.

$$\prod_{k=1}^{\infty} \frac{1 + (\pm 1)^k \alpha(k)}{1 - (\pm 1)^k \alpha(k)} > \prod_{k=1}^{\infty} (1 + (\pm 1)^k 2\alpha(k))$$

Położenie $(\pm 1)^k \alpha(k) > a > 0$ dla N różnych k

implikuje, że $(\pm 1)^m \rho(m) > 0$ zachodzi

dla co najmniej $(1 + 2a)^N$ różnych m , ponieważ $|\rho(k)| \leq 1$.