# Prawdopodobieństwo i statystyka

Wykład XIII:
Prognoza. Warunkowa wartość oczekiwana

26 stycznia 2015

# Zagadnienie prognozowania

- Przypuśćmy, że mamy dany ciąg liczb $x_1, x_2, \ldots, x_n$, stanowiących wyniki pomiaru pewnej zmiennej w czasie wielkości $x$, w chwilach $t_1 < t_2 < \ldots < t_n$. Inaczej mówiąc, mamy dany „szereg czasowy".
- Zagadnienie prognozowania: Niech $T > t_n$. Jaką wartość przyjmie badana wielkość w chwili $T$?
- Jeżeli $x$ jest funkcją tylko czasu $t$, tzn. $x_k = f(t_k), k = 1, 2, \ldots, t_n$, możemy próbować odgadnąć postać funkcji $f$, np.
  - znajdując współczynniki wielomianu interpolacyjnego,
  - lub amplitudę, częstość i przesunięcie sygnału sinusoidalnego,
  - lub parametry przekształcenia $S$, którego kolejne iteracje $S(t_0), S^2(t_0), \ldots S_n(t_0)$ dają nam kolejne wartości $x_1$, $x_2$, $\ldots$, $x_n$.
- To jest jednak rzadka sytuacja. Na ogół musimy zakładać, że liczby $x_1, x_2, \ldots, x_n$ są wartościami ciągu zmiennych losowych.

# Pojęcie prognozy (predykcji)

Postawienie zagadnienia: znamy rozkład łączny zmiennych losowych $X_1, X_2, \ldots, X_n$, $Y_1$, $Y_2, \ldots$, $Y_m$, gdzie $X_1, X_2, \ldots, X_n$ reprezentują „przeszłość", a $Y_1$, $Y_2, \ldots$, $Y_m$ - „przyszłość". Na podstawie przeszłości chcemy ocenić wartości przyszłe w postaci funkcji $f(X_1, X_2, \ldots, X_n)$. Jako „miarę jakości" prognozy przyjmujemy „błąd średniokwadratowy"

$$E\|(Y_1, Y_2, \ldots, Y_m)^T - f(X_1, X_2, \ldots, X_n)\|^2 = \sum_{i=1}^{m} E(Y_i - f_i(X_1, X_2, \ldots, X_n))^2.$$

Pokażemy, że błąd średniokwadratowy jest minimalizowany przez

$$f_i(X_1, X_2, \ldots, X_n) = E\left(Y_i \middle| (X_1, X_2, \ldots, X_n)\right), i = 1, 2, \ldots, m.$$

# Rozkłady warunkowe

Niech $\vec{Y}$ i $\vec{Z}$ będą wektorami losowymi o wartościach w $\mathbb{R}^m$ i $\mathbb{R}^n$, określonymi na tej samej przestrzeni probabilistycznej $(\Omega, \mathcal{F}, P)$. Jeżeli $P(\vec{Z} = \vec{z}) > 0$, to rozkładem warunkowym wektora $\vec{Y}$ gdy $\vec{Z} = \vec{z}$ nazywamy prawdopodobieństwo

$$\mathbb{R}^m \supset A \mapsto P_{\vec{Y}|\vec{Z}=\vec{z}}(A) = P(\vec{Y} \in A | \vec{Z} = \vec{z}) \left( = \frac{P(\vec{Y} \in A, \vec{Z} = \vec{z})}{P(\vec{Z} = \vec{z})} \right).$$

Pytanie: jak określić rozkład warunkowy w ogólnym przypadku? Jeżeli $P_{(Y,Z)}$ jest absolutnie ciągły z gęstością $p_{Y,Z}(y,z)$, to można określić gęstość rozkładu $P_{Y|Z=z}$ za pomocą wzoru

$$p_{Y|Z=z}(y) = \begin{cases} \frac{p_{Y,Z}(y,z)}{\int_{-\infty}^{+\infty} p_{Y,Z}(u,z)\, du}, & \text{jeśli } \int p_{Y,Z}(u,z)\, du > 0 \\ \mathbb{1}_{[0,1]}(y), & \text{jeśli } \int p_{Y,Z}(u,z)\, du = 0 \end{cases}.$$

# Definicja i własności warunkowej wartości oczekiwanej

Mając dany rozkład warunkowy $P_{\vec{Y}|\vec{Z}=\vec{z}}(\cdot)$ określamy

$$\begin{aligned} E(\vec{Y}|\vec{Z}=\vec{z}) &:= \int \vec{y} P_{\vec{Y}|\vec{Z}=\vec{z}}(d\vec{y}), \\ E(\vec{Y}|\vec{Z}) &:= E(\vec{Y}|\vec{Z}=(\cdot)) \circ \vec{Z}. \end{aligned}$$

- Jeżeli $E|U| < +\infty$ i $E|V| < +\infty$, to
$$E(\alpha U + \beta V|\vec{Z}) = \alpha E(U|\vec{Z}) + \beta E(V|\vec{Z}).$$
- Jeżeli $E|Y| < +\infty$ i $|h(\vec{Z})| \leqslant C$, to
$$E(h(\vec{Z}) \cdot Y|\vec{Z}) = h(\vec{Z}) \cdot E(Y|\vec{Z}).$$

# Warunkowa wartość oczekiwana jako rzut ortogonalny

- Jeżeli $EY^2 < +\infty$ i $E(h(\vec{Z}))^2 < +\infty$, to
$$E(h(\vec{Z}) \cdot Y|\vec{Z}) = h(\vec{Z}) \cdot E(Y|\vec{Z}).$$
- Jeżeli $EY^2 < \infty$, to
$$\mathrm{Var}\,(Y) = E(Y - E(Y|\vec{Z}))^2 + \mathrm{Var}\,(E(Y|\vec{Z})).$$

**Twierdzenie (Warunkowa wartość oczekiwana jako minimalizator)**

Niech $EY^2 < +\infty$ i $\vec{Z}$ przyjmuje wartości w $\mathbb{R}^n$. Wówczas $E(Y|\vec{Z})$ jest jedynym minimalizatorem funkcjonału
$$h \mapsto E(Y - h(\vec{Z}))^2,$$
gdy $h$ przebiega zbiór $\{h : \mathbb{R}^n \to \mathbb{R}^1\,;\ E(h(\vec{Z}))^2 < +\infty\}$.

# Warunkowa wartość oczekiwana jako rzut ortogonalny - cd.

Uwaga: w terminach przestrzeni Hilberta $L^2(\Omega, \mathcal{F}, P)$ warunkowa wartość oczekiwana jest rzutem ortogonalnym na podprzestrzeń funkcji postaci $\{h(\vec{Z})\}$, czyli funkcji $\sigma(\vec{Z})$-mierzalnych. W tym kontekście (niemal) oczywiste są następujące fakty:

- Jeżeli $E|Y| < +\infty$ i $g : \mathbb{R}^n \to \mathbb{R}^m$, to

$$E\left(E(Y|\vec{Z})|g(\vec{Z})\right) = E(Y|g(\vec{Z})).$$

- Jeżeli $\vec{Z}$ jest funkcją stałą, to $E(Y|\vec{Z}) = EY$.
- Co by było, gdybyśmy minimalizowali $E|Y - h(\vec{Z})|$?

# Prognoza liniowa

W zagadnieniu prognozy zmiennych $Y_1, Y_2, \ldots, Y_m$, na podstawie $X_1, X_2, \ldots, X_n$ poszukujemy najlepszego przybliżenia zmiennych $Y_i$ w postaci $f_i(X_1, X_2, \ldots, X_n)$, gdzie $f_i$ spełnia tylko ogólne warunki całkowalności, należy więc do bardzo szerokiej klasy funkcji.
Z prognozą liniową mamy do czynienia, gdy poszukujemy najlepszego przybliżenia w klasie funkcji

$$f_i(X_1, X_2, \ldots, X_n) = \sum_{j=1}^{n} \alpha_{i,j} X_j, \ \ i = 1, 2, \ldots, m.$$

To na ogół dużo łatwiejsze zadanie!

Uwaga: Istnieje ważna klasa szeregów czasowych, dla których oba pojęcia prognozy pokrywają się: są to procesy gaussowskie.

# Procesy gaussowskie

## Definicja zmiennych losowych gaussowskich

Mówimy, że zmienne losowe $X_1, X_2, \ldots, X_n$ są gaussowskie, jeśli ich dowolna kombinacja liniowa $\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n$ ma jednowymiarowy rozkład normalny, tzn.

$$\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n \sim \mathcal{N}(m_{\vec{\alpha}}, \sigma^2_{\vec{\alpha}}),$$

gdzie $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$.

Dopuszczamy przypadek $\sigma^2_{\vec{\alpha}} = 0$. Z definicji $\mathcal{N}(m, 0) = \delta_m$.

## Rodziny gaussowskie

Rodzinę zmiennych losowych $\{X_i\}_{i \in \mathbb{I}}$ nazywamy gaussowską, jeśli dla każdego skończonego podzbioru $\{i_1, i_2, \ldots, i_n\} \subset \mathbb{I}$ zmienne $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$ są gaussowskie.

# Procesy gaussowskie - cd.

Biorąc $\vec{\alpha} = (0, \dots, 0, 1, 0, \dots, 0)T$, otrzymujemy rozkład normalny dla składowych $X_k \sim \mathcal{N}(m_k, \sigma_k^2)$. W ogólności,

$$m_{\vec{\alpha}} = E(\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n) = E\langle \vec{\alpha}, \vec{X} \rangle = \langle \vec{\alpha}, E\vec{X} \rangle.$$

Podobnie

$$\sigma^2_{\vec{\alpha}} = \mathrm{Var}\,(\langle \vec{\alpha}, \vec{X} \rangle) = \langle \vec{\alpha}, \mathrm{Cov}\,(\vec{X})\,\vec{\alpha} \rangle.$$

## Twierdzenie (Transformacja liniowa zmiennych gaussowskich)

Jeżeli wektor losowy $\vec{X} = (X_1, X_2, \dots, X_n)^T$ ma składowe gaussowskie, przy czym $E\vec{X} = \vec{m}$ i $\mathrm{Cov}\,(X) = \Sigma)$ i jezeli $A : \mathbb{R}^n \to \mathbb{R}^m$ jest odwzorowaniem liniowym, to składowe wektora $A(\vec{X})$ też są gaussowskie, przy czym

$$EA(\vec{X}) = A(\vec{m}), \; \mathrm{Cov}\,(A(\vec{X})) = A\Sigma A^T.$$

# Konstrukcja zmiennych gaussowskich

## Twierdzenie (Konstrukcja zmiennych gaussowskich)

Jeżeli $\vec{m} \in \mathbb{R}^n$ i $\Sigma$ jest macierzą $n \times n$, symetryczną i nieujemnie określoną, to istnieje wektor losowy $\vec{X}$ o składowych gaussowskich, który spełnia związki

$$E\vec{X} = \vec{m}, \quad \mathrm{Cov}\,(\vec{X}) = \Sigma.$$

## Twierdzenie (Charakterystyka rozkładu łącznego zmiennych gaussowskich)

Rozkład łączny zmiennych losowych gaussowskich $(X_1, X_2, \ldots, X_n)$ (nazywany $n$-wymiarowym rozkładem normalnym) jest w pełni określony przez swoja wartość oczekiwaną $\vec{m}$ i macierz kowariancji $\Sigma$. Piszemy $\vec{X} \sim \mathcal{N}(\mathbf{m}, \Sigma)$.

# Niezależność zmiennych gaussowskich

## Twierdzenie (Absolutna ciągłość rozkładu normalnego)

Rozkład normalny jest absolutnie ciągły dokładnie wtedy, gdy macierz $\Sigma$ jest nieosobliwa ($\det(\Sigma) \neq 0$). W takim przypadku gęstość zadana jest wzorem:

$$p_{\vec{m},\Sigma}(\vec{x}) = \frac{1}{(\sqrt{2\pi})^d} \frac{1}{\sqrt{\det \Sigma}} \exp\left( - \frac{1}{2} \langle \vec{x} - \vec{m}, \Sigma^{-1} (\vec{x} - \vec{m}) \rangle \right).$$

## Twierdzenie (Niezależność zmiennych gaussowskich)

Zmienne gaussowskie $X_1, X_2, \ldots, X_n$ są niezależne dokładnie wtedy, gdy są nieskorelowane:

$$\operatorname{cov}(X_i, X_j) = 0, \;\; i, j = 1, 2, \ldots, n, \; i \neq j.$$

# Prognoza dla zmiennych gaussowskich

**Twierdzenie (Prognoza dla zmiennych gaussowskich)**

Jeżeli zmienne $X_1, X_2, \ldots, X_n$, $Y_1$, $Y_2, \ldots$, $Y_m$ są gaussowskie, to prognoza liniowa $Y_1$, $Y_2, \ldots$, $Y_m$ na podstawie $X_1, X_2, \ldots, X_n$ pokrywa się z pełną prognozą (tzn. przybliżeniem $Y_1$, $Y_2, \ldots$, $Y_m$ za pomocą zmiennych postaci $h(X_1, X_2, \ldots, X_n)$).

Uwaga: Jeżeli w schemacie prognozy

$$X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$$

zmienne są gaussowskie, to prognoza (liniowa) $Y_1$, $Y_2, \ldots$, $Y_m$ na podstawie $X_1, X_2, \ldots, X_n$ jest również wektorem o składowych gaussowskich. Jak określić wektor prognozy?