PSK – Informatyka st. stacj. & niestacj.

Temat Porównaj rozkładu i ich znaczenie.

Wstęp. Niech dane będzie ZL oraz jego MPK (Ω, Σ, P).
Z teoretycznego punktu widzenia wynikiem obserwacji ZL jest procedura

$$\Omega \ni \omega \longrightarrow X(\omega) \in \mathbb{R},$$

gdzie X oznacza zmienną losową stowarzyszoną z tą obserwacją.
Jeśli X jest dyskretna, to $\Omega = \dot{\bigcup\limits_{n \in \mathbb{N}_0}} A_n$, $\mathbb{N}_0 \subset \mathbb{N}$

oraz $\forall_{\omega \in A_n} X(\omega) = x_n \in \mathbb{R}$, gdzie

$P(A_n) = p_n \in (0,1)$. Zatem obserwujemy x_n nie na pewno.
Jeśli X jest ciągła, to $\forall_{x_0 \in \mathbb{R}} P(\{\omega \in \Omega : X(\omega) = x_0\}) = 0$,

nie dziwi bardziej $X(\omega)$ nie może być traktowane jako pomiar obserwacji $\omega \in \Omega$.
Mamy nie problem: co jest wynikiem pomiaru obserwacji ZL?

Rozwiązaniem tego problemu są właśnie parametry rozkładu.

Zaczniemy od przypadku dyskretnego.

I. Niech X ma rozkład dyskretny:

$$\Omega = \dot{\bigcup_{n\in N_0}} A_n, \quad N_0 \subset \mathbb{N}$$

$$X(\omega) = x_n, \quad \omega \in A_n, \quad P(A_n) = p_n \in (0,1]$$
$$\sum_{n\in N_0} p_n = 1.$$

Bierzemy ciąg liczb $(x_n p_n)_{n\in N_0}$. Jeśli jest on sumowalny, czyli $\sum_{n\in N_0} x_n p_n \in \mathbb{R}$, to tę liczbę oznaczamy EX, czyli

$$EX = \sum_{n\in N_0} x_n p_n$$

i nazywamy wartością oczekiwaną zmiennej losowej X albo rozkładu X.

Uwagi 1°. Jeśli X ma rozkład skończony, to EX istnieje zawsze.
Wtedy $|N_0| = k \geq 2$
$$EX = \sum_{i=1}^{k} x_i p_i.$$
Taką liczbę nazywamy średnią ważoną (z wagami p_i)

W szczególności, gdy X ma rozkład <u>jednorodny</u>,

$p_1 = \ldots = p_k = p$, to $p = \frac{1}{k}$ oraz

$$EX = \frac{1}{k}\sum_{i=1}^{k} x_i \text{ – średnia arytmetyczna.}$$

Dlatego EX nazywamy też <u>średnią</u> i oznaczamy

przez m („mean" z ang.).

2^0. Nie każdy rozkład X ma <u>średnią</u>.

3^0. Operacja

$$X \longrightarrow EX \in \mathbb{R}$$

ma własność LINIOWOŚCI, tzn.

jeśli X, Y mają średnią, to dla $a, b \in \mathbb{R}$

$aX + bY$ też ma oraz

$$E(aX+bY) = a(EX) + b(EY).$$

4^0. Jeśli $m = EX$, to zmienna losowa

$X - m$ ma średnią $= 0$.

Nazywamy to „centrowaniem" rozkładu.

5°. Weźmy rozkład X oraz nowy

$$Y = X^2.$$

Jeśli Y ma średnią, to mówimy, że X ma <u>drugi moment</u> i oznaczamy m_2.

Wtedy $$EY = E(X^2) = \sum_{n \in \mathbb{N}_0} x_n^2 p_n =: m_2.$$

6°. Jeśli istnieje m_2, to istnieje m.

<u>Przykłady</u>

1°. X :

0	1
q	p

$m = EX = p$

2°. $X \in B(n,p)$. Z zasady faktoryzacji

$$X = X_1 + X_2 + \dots + X_n,$$ gdzie

każda zm. X_j ma rozkład jak w P1.

Zatem z liniowości operacji E

$$m = EX = EX_1 + EX_2 + \dots + EX_n = \underbrace{p + p + \dots + p}_{n} = np$$

Czyli
$$\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}\cdot k = np$$

3°. $X \in \mathcal{P}(\lambda)$. Wtedy
$$\sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}$$
$$= e^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\cdot\lambda\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}.$$

Ale $\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = e^{\lambda}$.

Zatem $\boxed{m = EX = \lambda}$.

4°. X ma rozkład geometryczny.

Oblicz EX.

<u>Wariancja jako drugi moment centralny</u>.

Niech X ma m_2. Zatem X ma m.
Bierzemy $X-m$, oraz kwadrat $(X-m)^2$.
Wtedy wartość oczekiwaną $(X-m)^2$ nazywamy

<u>wariancją</u> zm. losowej X i piszemy

$$\mathrm{var}(X) = E(X-m)^2.$$

<u>Uwagi</u>

1°. $\mathrm{var}(X)$ nazywamy też „średnim kwadratowym odchyleniem"

i oznaczamy σ^2 $(\sigma \geq 0)$.

2°. $\mathrm{var}(X) \geq 0$.

3°. $\mathrm{var}(X) = 0 \equiv X = \mathrm{const}$.

4°. $\sigma = \sqrt{\mathrm{var}(X)}$ - nazywamy <u>dyspersją</u> (albo rozproszeniem) rozkładu.

5°. Jeśli X ma m_2, to

$$\mathrm{var}(X) = m_2 - m^2 \left(= E(X^2) - (EX)^2 \right).$$

6°. Ponieważ sytuacja $X = \mathrm{const}$ nie zdarza się w przyrodzie, <u>każda zjawisko losowe</u> X ma <u>wariancję dodatnią</u>!

7°. Jeśli X, Y mają 2-ie momenty i są stochastycznie niezależne, to
$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$$
oraz $\mathrm{var}(aX) = a^2\,\mathrm{var}(X)$, dla $a \in \mathbb{R}$.

<u>Przykłady</u> 1°. X

0	1
q	p

$m = m_2 = p$, więc
$$\sigma^2 = \mathrm{var}(X) = p - p^2 = p(1-p) = p \cdot q$$

2°. $X \in B(n, p)$.

Korzystamy z faktu
$$X = X_1 + \ldots + X_n.$$ Wtedy X_j mają rozkład jak w 1° i są stochast. niezależne.

Stąd $$\mathrm{var}(X) = \underbrace{pq + \ldots + pq}_{n} = n \cdot p \cdot q$$

3°. $X \in P(\lambda)$
$$\mathrm{var}(X) = ?$$ obliczyć!

II Przypadek rozkładu ciągłego.

Niech X jak wyżej. Jeśli istnieje całka

$$\int_{-\infty}^{+\infty} t f_X(t)\,dt \text{, gdzie } f_X \text{ - f. gęstości } X,$$

to oznaczamy ją EX i nazywamy wart. oczekiw. lub średnią. Mamy

$$m = EX = \int_{-\infty}^{+\infty} t f_X(t)\,dt.$$

Operacja $X \longrightarrow EX$ (jak w przypadku dyskretnym) jest liniowa.

W szczególności $E(X-m) = 0$ - centrowanie.

Przykłady. $X \in \mathcal{J}([a,b])$.

Wtedy $$\int_{-\infty}^{+\infty} t f_X(t)\,dt = \frac{1}{b-a}\int_a^b t\,dt = \frac{1}{b-a}\,\frac{1}{2}t^2\Big]_a^b = \frac{b^2-a^2}{2(b-a)}$$

Zatem $EX = \frac{a+b}{2}$ - środek przedziału $[a,b]$

2° $X \in N(m, \sigma^2)$.

Wtedy $EX = m$

3°. $X \in N(\lambda)$, to $EX = \frac{1}{\lambda}$

Podobnie jak w przypadku dyskretnym, biorąc X^2, jeśli X^2 ma wartość oczekiwaną, to mówimy, że X ma 2-gi moment i piszemy

$$m_2 = E(X^2) = \int_{-\infty}^{+\infty} t^2 f_X(t)\,dt.$$

W szczególności, 2-gi moment <u>centralny</u> oznaczamy $var(X)$ i mamy

$$\sigma^2 = var(X) = E(X-m)^2 = E(X^2) - (E(X))^2$$
$$= m_2 - m^2,$$

gdzie $\sigma \geq 0$.

<u>Przykłady</u>

1°. $X \in J([a,b])$, $\sigma^2 = \frac{(b-a)^2}{12}$,

2^0. $X \in N(m, \sigma^2)$

$Var(X) = \sigma^2$.

Wtedy $\frac{X-m}{\sigma} \in N(0,1)$

$\hookrightarrow$ zasada standaryzacji

3^0 $X \in W(\lambda)$

$\sigma^2 = Var(X) = \frac{1}{\lambda^2}$

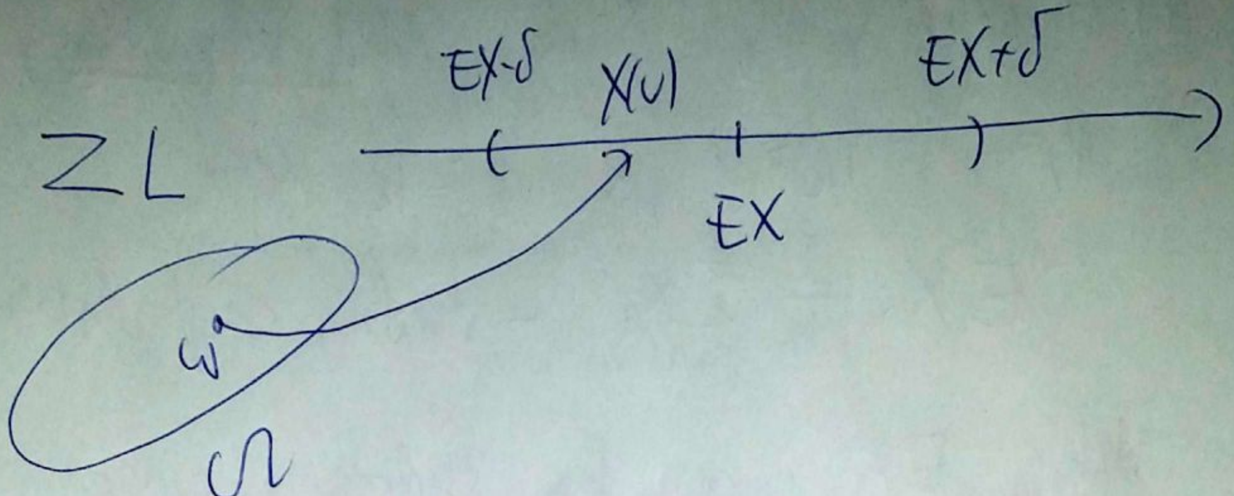
Zasada pomiaru i jego niepewności

Mamy ZL o rozkładzie $X \neq const$ z drugim momentem. Jak wiemy, wtedy $\sigma^2 > 0$ oraz istnieje $m = EX$.

Dla $\delta > 0$ weźmy zdarzenie

$$A_\delta = \{\omega \in \Omega : |X(\omega) - EX| < \delta\},$$

które śledzi zjawisko, że wynik (teoretyczny) obserwacji $X(\omega)$ należy do przedziału $(EX - \delta, EX + \delta)$



Można udowodnić, że

$$\forall_{\varepsilon>0} \exists_{\delta>0} \; P(A_\delta) \geq 1-\varepsilon, \text{ czyli}$$

1°. Z dowolnie dużym prawdop. $(=1-\varepsilon)$ zdarzy się, że teoretyczna wartość obserwacji $X(\omega)$ będzie z przedziału o środku EX i długości 2δ.

2°. $\delta \approx \sqrt{Var(X)} = \sigma$

zatem ~~rozproszenie~~ σ jest miarą <u>różnicy</u> pomiędzy EX a $X(\omega)$, czyli <u>miarą rozproszenia</u>.

3°. Wtedy $P(A_\delta)$ jest <u>miarą niepewności</u>

pomiaru, w którym przyjmujemy, że wynikiem jest $EX = m$ – wartość średnia (a nie X_i!)

Problem:

Obserwując ZL musimy przyjąć, że na tym etapie nie znamy rozkładu X.
Z drugiej strony, za wynik obserwacji bierzemy $m = EX$ i kontrolujemy go rozproszeniem σ.
Jak pozyskać m i σ nie znając rozkładu?

Odpowiedzią na to pytanie są dwa fundamentalne twierdzenia TEORII PRAWDOP. zwane TWIERDZENIAMI GRANICZNYMI.

I Tw. Graniczne – MOCNE PRAWO WIELKICH LICZB BERNOULLIEGO

Twierdzenie to rozwiązuje problem wyznaczenia $m = EX$ w sytuacji, kiedy nie możemy wyliczyć tej liczby (bo np. nie znamy rozkładu X).

ZAŁOŻENIA (MPWLB)

1°. X ma co najmniej 2-i moment

2°. (X_n) jest ciągiem niezależnych zmiennych losowych każda o tym samym rozkładzie co X.

Bierzemy $A = \{\omega \in \Omega : \frac{1}{n}(X_1(\omega) + \ldots + X_n(\omega)) \longrightarrow m\}$

Wtedy $\underline{P(A) = 1}$.

<u>Uwagi</u>. 1°. Praktyczne zastosowanie tw. polega na tym, że dla odpowiednio dużych n, $\omega_0 \in \Omega$

$$\frac{1}{n}(X_1(\omega_0) + X_2(\omega_0) + \ldots + X_n(\omega_0)) \approx m$$

2°. Wiemy, że $m = EX$ jest uważane jako wynik obserwacji ZL. Zatem MPWL jest dowodem na to, co jest obserwowane na co dzień:

Poznanie „części" nowego (ZL) wymaga metodyki obserwacji, po to aby wyniki tych obserwacji po ich uśrednieniu uważać za pomiar ZL.

3°. Weźmy rozkład $X: \begin{array}{c|c} 0 & 1 \\ \hline q & p \end{array}$

Aby wyznaczyć p, zgodnie z MPWLB bierzemy $(X_n)_{n\geq 1}$ - kopie rozkładu X i wtedy dla $\omega_0 \in \Omega$ (dane obserwacji)

$$\frac{1}{n}(X_1(\omega_0) + \ldots + X_n(\omega_0)) \approx EX = p$$

Ale lewa strona

$$f_n(\omega_0) = \frac{\#\text{obserwacji pojawienia się „1"}}{\#\text{powtórzeń}}$$

to frakcja zaobserwowanych # sukcesów („1") w serii powtórzeń, czyli

$$f_n(\omega_0) \cong p$$

prawd. TEORETYCZNE

prawd. EMPIRYCZNE

II Tw. graniczne nazywa się CENTRALNYM TW. GRANICZNYM

CTG opisuje ewolucję rozkładu, jeśli podlega on procedurze uśredniania.

ZAŁOŻENIA. X ma co najmniej drugi moment, a więc istnieje $m = EX$ i $\sigma^2 = var(X) > 0$.

Bierzemy ciąg niezależnych kopii X, czyli $(X_n)_{n \geq 1}$.

1) Uśredniamy go

$$Y_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

2) standaryzujemy

$$Z_n = \frac{Y_n - EY_n}{\sqrt{var(Y_n)}} = \frac{Y_n - m}{\frac{\sigma}{\sqrt{n}}}$$

Wtedy

$$F_{Z_n}(t) = P(\{\omega \in \Omega : Z_n(\omega) < t\}) \longrightarrow \Phi(t)$$

(dystr. $N(0,1)$).

Uwagi

1°. Praktyczne zastosowanie CTG

Dla dużych n mamy przybliżenie

$$P\left(\left\{\omega\in\Omega: \frac{\frac{1}{n}(X_1(\omega)+\dots+X_n(\omega))-m}{\frac{\sigma}{\sqrt{n}}}<t\right\}\right)\simeq\Phi(t)$$

2°. Wersja klasyczna: Tw. Moivre'a-Laplace'a jako przykład aproksymacji rozkładu $B(n,p)$.

Niech $B_n\in B(n,p)$, $n\geq 2$.

Oszacujmy $P(\{\omega\in\Omega: a\leq B_n(\omega)<b\})=$

$$=P(\{\omega\in\Omega: a\leq X_1(\omega)+\dots+X_n(\omega)<b\})$$

na mocy faktoryzacji B_n. Dalej standaryzujemy

$$=P\left(\left\{\omega\in\Omega: \frac{a}{n}\leq\frac{1}{n}(X_1(\omega)+\dots+X_n(\omega))<\frac{b}{n}\right\}\right)=$$

$$=P\left(\left\{\omega\in\Omega: \frac{\frac{a}{n}-p}{\sqrt{\frac{pq}{n}}}\leq\frac{\frac{1}{n}(X_1(\omega)+\dots+X_n(\omega))-p}{\sqrt{\frac{pq}{n}}}<\frac{\frac{b}{n}-p}{\sqrt{\frac{pq}{n}}}\right\}\right)$$

$$\stackrel{\text{C.T.G.}}{\cong}\Phi\left(\frac{\frac{b}{n}-p}{\sqrt{\frac{pq}{n}}}\right)-\Phi\left(\frac{\frac{a}{n}-p}{\sqrt{\frac{pq}{n}}}\right).$$