Statystyka Opisowa z Demografią oraz Biostatystyka

Zmienne losowe

Aleksander Denisiuk

denisjuk@euh-e.edu.pl

Elbląska Uczelnia Humanistyczno-Ekonomiczna

ul. Lotnicza 2

82-300 Elbląg

Zmienne losowe

Najnowsza wersja tego dokumentu dostępna jest pod adresem

`http://denisjuk.euh-e.edu.pl/`

Zmienne losowe

- w wyniku doświadczenia przyjmuje wartości z pewnego zbioru lizb rzeczywistych z określonym prawdopodobeństwem.
- przykłady
  - liczba przedmiotow, wyprodukowanych na danych stanowisku w ciągu zmiany
  - ilość energii elektrycznej, zużywanej przez dziennie przez zakład
  - czas rozmowy telefonicznej
  - wartość cech jednostek statystycznych, wylosowanych z populacji generalnej

Odwzorowanie na pozdziór liczb rzeczywistych

- każdy zbiór zdarzeń elementarnych można odwzorować na podzbiór liczb rzeczywistych
- na przykład:
  - rzucamy trzy razy monetami
  - rzucamy kostką

Definicja zmiennej losowej

- Dana jest przestrzeń probabilistyczna (E, S, P)
- *zmienna losowa* — to funkcja mierzalna względem S, określona na E, przybierająca wartości ze zbioru liczb rzeczywistych
  - zmienne losowe: X, Y, Z
  - wartości (*realizacje*): x, y, z
  - przyporządkowanie *zradzenie* $\mapsto$ *wartość zmiennej losowej* jest jednoznacznym
- zmienne losowe
  - skokowe (dyskretne)
  - ciągłe

Zmienne skokowe

- Funkcja rozkładu prawdopodobieństwa (rozkład prawdopodobieństwa): $P(X = x_i) = p_i$
  - $p_i - \geqslant 0$
  - $\sum_{i=1}^{\infty} p_i = 1$
- liczby x_i — punkty skokowe
- prawdopodobieństwa p_i — skoki
- przykład:
  - liczba orłów w trzech rzutach monetą

Dystrybuanta

- $F(x) = P(X < x)$, $x \in \mathbb{R}$
- dla zmiennej skokowej:
  - $F(x) = \sum_{x_i < x} p_i$
- przykład:
  - liczba orłów w trzech rzutach monetą
- znając rozkład, zawsze można znaleźć dystrybuantę
- znając dystrybuantę, zawsze można znaleźć rozkład

Właściwości dystrybuanty

- $0 \leqslant F(x) \leqslant 1$
- jest funkcją niemalejącą
- jest funkcją lewostronnie ciągłą
- $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$
- $P(a \leqslant x \leqslant b) = F(b) - F(a)$

Wartość oczekiwana zmiennej losowej

- $E(x) = \sum_{i=1}^{\infty} x_i p_i$ (jeżeli szereg jest bezwzględnie zbieżny)
  - $E(C) = C$ ($C =$ const)
  - $E(X + Y) = E(X) + E(Y)$
  - $E(X \cdot Y) = E(X) \cdot E(Y)$ dla *niezależnych* zmiennych losowych X i Y
  - $E(C \cdot Y) = C \cdot E(Y)$ ($C =$ const)

Wariancja zmiennej losowej

- $D^2(x) = \sum_{i=1}^{\infty}(x_i - E(X))^2 p_i = E(X - E(X))^2 = E(X^2) - \big(E(X)\big)^2$
  - $D^2(C) = 0$ ($C =$ const)
  - $D^2(C \cdot Y) = C^2 \cdot D^2(Y)$ ($C =$ const)
  - $D^2(X \pm Y) = D^2(X) + D^2(Y)$ dla *niezależnych* zmiennych losowych X i Y
- $D(X) = \sqrt{D^2(X)}$ — odchylenie standardowe

Przykłady zmiennych losowych

- W tramwaju zgasło światło w momencie, gdy jeden z pasażerów szukał biletu w celu skasowania. Pasażer miał 5 biletów po $2{,}4$ zł, 3 po $1, 2$ zł oraz 2 po $1{,}6$ zł. Jaka jest wartość oczekiwana i odchylenie standardowe?
- Średnia oczekiwana wygranej w Multi Multi to 1 zł

Rozkład dwumianowy

- dotyczy wielokrotnej realizacji doświadczenia, w wyniku którego możliwe są dwa zdarzenia: A (sukces) lub $\bar{A}$ — niepowodzenie. Zmienna losowa — ilość sukcesów
- $P(A) = p$, $P(B) = q = 1 - p$
- przykład — rzut monetą (niesymetryczną)
- dwókrotne doświadczenie
- trzykrotne doświaczenie
- n-krotne: $P(X = k) = \binom{n}{k} p^k q^{1-k}$
- $E(X) = np$, $D^2(X) = npq$

Rozkład dwumianowy. Przykłady

- gabinet kinezyterapii jest wyposażony w pięć wiaderek TummyTub. Na podstawie obserwacji obliczono, że prawdopodobieństwo tego, że jedno wiaderko jest wolne wynosi $0{,}1$. Oblicz prawdopodobieństwe tego, że wolne są dwa wiaderka, przynajmniej dwa wiaderka
- rzucamy pięć razy monetą symetryczną. Jakie jest prawdopodobieństwo trzykrotnego wyrzucenia orła?

Rozkład Possona

- rozkłąd rzadkich zdarzeń
- $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$
- jest przybliżeniem rozkładu dwumianowego, gdy n jest duże ($n>100$), zaś p jest małe ($p<0{,}2$), wówczas $\lambda=np$
- $E(X)=D(X)=\lambda$
- przykład:
  - 90 studentów, rejestracja nieobecności

liczba dni	0	1	2	3	4	5	6	7
liczba studentów	12	20	27	18	7	3	6	1

  - zakładamy rozkład Poissona
    - znaleźć dystrybuantę
    - prawdopodobieństwo nieobecności mniej niż dwa dni

Zmienne losowe ciągłe

- funkcja gęstości $f(x)$
- $P(a \leqslant X \leqslant b) = \int\limits_a^b f(x)\,dx$
- $P(x = a) = 0$
- dystrybuanta $F(x) = P(X < x) = \int\limits_{-\infty}^{x} f(u)\,du$
- $f(x) = \frac{dF}{dx} = F'(x)$
- $P(X < a) = F(a)$, $P(a < X < b) = F(b) - F(a)$,
  $P(X > a) = 1 - F(a)$
- $E(X) = \int\limits_{-\infty}^{\infty} x f(x)\,dx$
- $D^2(X) = \int\limits_{-\infty}^{\infty} \big(X - E(X)\big)^2 f(x)\,dx$

Rozkład Gaussa

- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$
- $N(m, \sigma)$
- $E(X) = m$, $D^2(X) = \sigma^2$
- $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt$
- $Z = \frac{X-m}{\sigma} \Longrightarrow N(0,1)$
  - $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
  - $F(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt = \Phi(x)$
    - czasami $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{x} e^{-\frac{t^2}{2}}\, dt$

Rozkład Gaussa. Przykłady

- Oblicz prawdopodobieństwo tego, że wzrost losowo wybranego męszczyzny będzie zawarty między 190 cm a 200 cm, $N(172, 6)$
- Oblicz $P(0 < X < 2)$, $P(X > 2)$, $P(X < -0{,}5)$, $P(|X| < 1)$ dla $N(0, 1)$