Kwantyle

- Kwantyl rzędu *p* rozkładu prawdopodobieństwa to taka liczba x_p, że

$$P\left(X \leq x_p\right) \geq p$$

$$P\left(X \geq x_p\right) \geq 1 - p$$

- Możemy go obliczyć z dystrybuanty:
  - Jeżeli F(x_p) = p, to x_p jest kwantylem rzędu p
  - Jeżeli F(x_p) > p oraz $\lim_{x \to x_p^-} F\left(x\right) < p$ to jak wyżej

Jak to policzyć w Pythonie?

- Bardzo prosto.

Jak to policzyć w Pythonie?

- Bardzo prosto.
- Przykład: dla rozkładu Poissona z parametrem $\lambda=2$ obliczamy kwantyl rzędu *p*:
  - Stworzenie rozkładu w Pythonie:

    ```
    rozklad = scipy.stats.poisson(2)
    ```

  - Obliczenie kwantyla rzędu *p*:

    ```
    rozklad.ppf(p)
    ```

  - Lub w jednej linijce jeśli ktoś bardzo chce

```
scipy.stats.poisson(2).ppf(p)
```

Jak to policzyć inaczej?

- W niektórych sytuacjach możemy nie znać dokładnej postaci rozkładu, a mieć tylko generator liczb losowych z danego rozkładu.

Jak to policzyć inaczej?

- W niektórych sytuacjach możemy nie znać dokładnej postaci rozkładu, a mieć tylko generator liczb losowych z danego rozkładu.
- W takim przypadku robimy symulację!
  - Losujemy N liczb losowych z rozkładu (X_i)
  - Sortujemy (numpy.sort) wylosowane wartości
  - Kwantylem rzędu p jest liczba o indeksie $N{\cdot}p$ lub średnia z elementów położonych najbliżej jeśli to nie liczba całkowita

Wartość oczekiwana

- Wartością oczekiwaną zmiennej losowej nazywamy
  - Dla zmiennej losowej dyskretnej

$$E[X]=\sum_{x\in\mathbb{R}} x\, pmf_X(x)=\sum_i x_i\, p_i$$

  - Dla zmiennej losowej ciągłej

$$E[X]=\int_{x\in\mathbb{R}} x\, f_X(x)\, dx$$

Momenty

- *k*-ty moment zwykły $m_k = E[X^k]$
  - dla k = 1 wartość oczekiwana

$$m_k = \sum_{x \in \mathbb{R}} x^k \, pmf_X(x) = \sum_i x_i^k \, p_i$$

$$m_k = \int_{x \in \mathbb{R}} x^k f_X(x) \, dx$$

- *k*-ty moment centralny $\mu_k = E[(X - E[X])^k]$
  - dla k = 1 zawsze równy 0
  - dla k = 2 wariancja

$$\mu_k = \sum_{x \in \mathbb{R}} (x - E[X])^k \, pmf_X(x) = \sum_i (x_i - E[X])^k \, p_i$$

$$\mu_k = \int_{x \in \mathbb{R}} (x - E[X])^k f_X(x) \, dx$$

Przykład

- Dyskretna zmienna losowa określona jest przez następującą funkcję (masy) prawdopodobieństwa

wartość	prawdopodobieństwo (pmf)
0	0.2
1	0.3
2	0.3
3	0.2

- Ile wynosi wartość oczekiwana i wariancja tego rozkładu prawdopodobieństwa?

Przykład – rozkład jednostajny

- Ciągła zmienna losowa określona jest przez następującą funkcję gęstości prawdopodobieństwa:

$$f_X(x) = \frac{1}{b-a}$$

  oraz 0 poza przedziałem [a; b]

- Ile wynosi wartość oczekiwana i wariancja tego rozkładu prawdopodobieństwa?

Przykład – rozkład wykładniczy

- Ciągła zmienna losowa określona jest przez następującą funkcję gęstości prawdopodobieństwa:

$$f_X(x) = \lambda e^{-\lambda x}$$

- Ile wynosi wartość oczekiwana i wariancja tego rozkładu prawdopodobieństwa?

Jak to policzyć w Pythonie?

- Bardzo prosto.

Jak to policzyć w Pythonie?

- Bardzo prosto.
- Przykład: dla rozkładu Poissona z parametrem $\lambda=2$ obliczamy średnią i wariancję:
  - Stworzenie rozkładu w Pythonie:

    `rozklad = scipy.stats.poisson(2)`
  - Obliczenie n-tego momentu zwykłego:

    `rozklad.moment(1)` lub `rozklad.mean()` dla n=1
  - Obliczenie wariancji:

    `rozklad.stats('v')` lub jako $E[X^2] - E[X]^2$
  - Lub w jednej linijce jeśli ktoś bardzo chce

```
scipy.stats.poisson(2).stats('v')
```

Jak to policzyć inaczej?

- W niektórych sytuacjach możemy nie znać dokładnej postaci rozkładu, a mieć tylko generator liczb losowych z danego rozkładu.

Jak to policzyć inaczej?

- W niektórych sytuacjach możemy nie znać dokładnej postaci rozkładu, a mieć tylko generator liczb losowych z danego rozkładu.
- W takim przypadku robimy symulację!
  - Losujemy N liczb losowych z rozkładu (X_i)
  - Szukana wartość oczekiwana
    = średnia z wylosowanych liczb

Jak to policzyć inaczej?

- Wartość oczekiwana

$$E[X] \approx \frac{1}{N} \sum_i X_i$$

- Wariancja

$$D^2[X] = E[X^2] - E[X]^2 = \frac{1}{N} \sum_i X_i^2 - \left(\frac{1}{N} \sum_i X_i \right)^2$$

- n-ty moment zwykły

$$E[X^n] = \frac{1}{N} \sum_i (X_i)^n$$

- n-ty moment centralny

$$E[(X - E[X])^n] = \frac{1}{N} \sum_i (X_i - E[X])^n$$

Jak to policzyć inaczej?

- Przykład: dla rozkładu Poissona z parametrem λ=2 obliczamy średnią i wariancję
  - Stworzenie rozkładu w Pythonie:

    ```
    rozklad = scipy.stats.poisson(2)
    ```

  - Wygenerowanie N liczb losowych:

    ```
    X = rozklad.rvs(size = N)
    ```

  - Średnia (pierwszy moment zwykły):

    ```
    μ = numpy.sum(X) / N
    ```

  - Wariancja (drugi moment centralny):

    ```
    numpy.sum((X – μ)**2) / N
    ```

Jak to policzyć inaczej?

- Można jeszcze prościej:
  - Średnia:

    ```
    numpy.mean(X)
    ```
  - Wariancja

    ```
    numpy.var(X)
    ```
  - n-ty moment centralny

    ```
    scipy.stats.moment(X, n)
    ```

Zadanie

- Dla wskazanego rozkładu, proszę:
  - Narysować funkcję gęstości prawdopodobieństwa
  - Obliczyć dla danego rozkładu średnią, wariancję i trzeci moment zwykły, dwiema metodami:
    - Korzystając z własności rozkładu (slajd nr 7)
    - Na podstawie wylosowanych wartości (`rvs` + …)

Zadanie

- rozkład normalny
  `norm(5, 2)`
- rozkład wykładniczy
  `expon(3)`
- rozkład Poissona
  `poisson(2)`
- rozkład jednorodny
  `uniform(2, 2)`
- rozkład Pareto
  `pareto(1)`
- rozkład gamma
  `gamma(5)`