# Typowy problem

- Chcemy dowiedzieć się, ile dla zadanej zmiennej losowej *X* i przedziału (a,b] wynosi
  - $P(a < X \leq b)$ czyli $P(X \in [a,b])$
- Wiemy, że dla zmiennej losowej
  - ciągłej, możemy obliczyć

$$P(a < X \leq b) = \int_a^b f_X(x)\,dx$$

  - dyskretnej, możemy obliczyć

$$P(a < X \leq b) = \sum_{x \in (a,b]} pmf_X(x)$$

# Typowy problem

- Chcemy dowiedzieć się, ile dla zadanej zmiennej losowej *X* i przedziału (a,b] wynosi
  - P(*a* < *X* ≤ *b*) czyli P( *X*∈[*a*,*b*) )
- Wiemy, że dla zmiennej losowej
  - ciągłej, możemy obliczyć

$$P(a < X \leq b) = \int_a^b f_X(x)\,dx$$

  - dyskretnej, możemy obliczyć

$$P(a < X \leq b) = \sum_{x \in (a,b]} pmf_X(x)$$

  - ale nie jest to najwygodniejszy sposób.

# Typowy problem

- Zauważmy, że z definicji dystrybuanty

$$F_X(x_0) = P(X \leq x_0)$$

# Typowy problem

- Zauważmy, że z definicji dystrybuanty

$$F_X(x_0) = P(X \leq x_0)$$

- Zatem, jeśli zauważymy, że

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

# Typowy problem

- Zauważmy, że z definicji dystrybuanty

$$F_X(x_0) = P(X \leq x_0)$$

- Zatem, jeśli zauważymy, że

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

to jednocześnie

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$$

# Typowy problem

- Zapis

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F_X(b) - F_X(a)$$

- Możemy uogólnić (nie do końca formalnie) na $\pm\infty$

$$P(X > c) = P(c < X \le \infty) = F_X(\infty) - F_X(c) =$$
$$= 1 - F_X(c)$$

$$P(X \le c) = P(-\infty < X \le c) = F_X(c) - F_X(-\infty) =$$
$$= F_X(c)$$

# Typowy problem

- Co, jeśli mamy inny przedział, np. P(a ≤ X < b)?
  - Dla zmiennych losowych <u>ciągłych</u> nie ma różnicy

$$P(a < X \le b) = P(a \le X < b) = P(a \le X \le b) = P(a < X < b) =$$
$$F_X(b) - F_X(a)$$

  - Dla zmiennych losowych, które nie są ciągłe:

$$P(a \le X \le b) = P(a < X \le b) + P(X = a) =$$
$$F_X(b) - F_X(a) + pmf_X(a)$$

$$P(a < X < b) = P(a < X \le b) - P(X = b) =$$
$$F_X(b) - F_X(a) - pmf_X(b)$$

$$P(a \le X < b) = P(a < X \le b) + P(X = a) - P(X = b) =$$
$$F_X(b) - F_X(a) + pmf_X(a) - pmf_X(b)$$

# Jak to policzyć w Pythonie?

- Bardzo prosto.

# Jak to policzyć w Pythonie?

- Bardzo prosto.
- Przykład: dla rozkładu Poissona z parametrem λ=2 obliczamy P(0 < *X* ≤ 3):
  - Stworzenie rozkładu w Pythonie:

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

  - Obliczenie ze wzoru:

    ```
    rozklad.cdf(3) − rozklad.cdf(0)
    ```

  - Lub w jednej linijce jeśli ktoś bardzo chce

```
scipy.stats.poisson(2).cdf(3) – scipy.stats.poisson(2).cdf(0)
```

# Jak to policzyć w Pythonie?

- Bardzo prosto.
- Przykład: dla rozkładu Poissona z parametrem λ=2 obliczamy $P(0 \leq X \leq 3)$:
  - Stworzenie rozkładu w Pythonie:

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

  - Obliczenie ze wzoru:

    ```
    rozklad.cdf(3) − rozklad.cdf(0) + rozklad.pmf(0)
    ```

  - Lub w jednej (długiej) linijce jeśli ktoś bardzo chce

```
scipy.stats.poisson(2).cdf(3) – scipy.stats.poisson(2).cdf(0)
  + scipy.stats.poisson(2).pmf(0)
```

# Jak to policzyć inaczej?

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

# Jak to policzyć inaczej?

- W niektórych sytuacjach możemy nie znać dystrybuanty, mieć tylko generator liczb losowych z danego rozkładu.
- W takim przypadku robimy symulację!
  - Losujemy N liczb losowych z rozkładu
  - Sprawdzamy, ile (K) z tych liczb wpadnie do naszego przedziału
  - Szukane prawdopodobieństwo ≈ K / N

# Jak to policzyć inaczej?

- Przykład: dla rozkładu Poissona z parametrem $\lambda=2$ obliczamy $P(0 \leq X \leq 5)$:
  - Stworzenie rozkładu w Pythonie:

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

  - Wygenerowanie N liczb losowych:

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

  - Sprawdzenie, ile z tych liczb należy do przedziału:

    ```
    K = numpy.sum(numpy.logical_and(0 <= x, x <= 5))
    ```

  - Obliczenie przybliżonego prawdopodobieństwa:

    ```
    float(K) / float(N)
    ```

# Zadanie

- Dla wskazanego rozkładu, proszę:
  - Narysować funkcję gęstości prawdopodobieństwa
  - Obliczyć prawdopodobieństwo przyjmowania przez zmienną losową o zadanym rozkładzie wartości ze wskazanego przedziału, dwiema metodami:
    - Na podstawie wartości dystrybuanty (`cdf`)
    - Na podstawie wylosowanych wartości (`rvs`)

# Zadanie

- rozkład normalny
  `normal(5, 2)`
  $P(0 \leq X \leq 3)$
- rozkład wykładniczy
  `expon(3)`
  $P(2 \leq X < 3)$
- rozkład Poissona
  `poisson(2)`
  $P(X \geq 2)$
- rozkład jednorodny
  `uniform(2, 2)`
  $P(X < 3)$
- rozkład Pareto
  `pareto(1)`
  $P(X > 10)$
- rozkład gamma
  `gamma(5)`
  $P(2 < X \leq 5)$