Дискретная модель

Определение

Дискретное вероятностное пространство

— упорядоченная тройка:

- Ω - пространство элементарных исходов
- $\mathcal{F} = 2^\Omega$ - множество всех допустимых событий
- $P : \mathcal{F} \to [0; 1]$ - вероятность $P(A) = \sum_{\omega \in A} P(\omega)$, где $P(\omega)$ - функция $\omega \to \mathbb{R}_{\geq 0}$, такая что $\sum_{\omega \in \Omega} P(\omega) = 1$

Лемма

Свойства вероятности в дискретном вероятностном пространстве

1. $P(A) \geq 0$ для любого события A
2. $P(\Omega) = 1$
3. $P(A_1 \sqcup A_2 \sqcup \ldots \sqcup A_m) = P(A_1) + \ldots + P(A_m)$
4. $P(A_1 \sqcup A_2 \sqcup \ldots) = P(A_1) + P(A_2) + \ldots$ - для счетного числа событий

Математическое ожидание в дискретном пространстве

Определение

Математическое ожидание дискретной случайной величины

$$EX = \sum_{x \in \mathrm{supp}(X)} x \cdot \mathrm{PMF}_X(x)$$

— среднее значение с учётом вероятностей.

PMF

Определение

> **PMF или функция вероятности в точке (probability mass function)**
>
> $$PMF_X(k) = P(X = k)$$

Лемма

> **Свойства PMF в дискретном пространстве**
>
> $$\begin{aligned} &\mathrm{PMF}_X(k) \geq 0 \quad \text{для всех } k \in \mathbb{R}; \\ &\sum_{x \in \mathrm{supp}(X)} \mathrm{PMF}_X(x) = 1. \end{aligned}$$

Определение

> **Дискретное распределение**
>
> — распределение, заданное с помощью PMF, удовлетворяющей вышеуказанным свойствам

Определение

> **Параметр распределения**
>
> — параметр PMF, такой, что свойства PMF выполнены при любом значении параметра из некоторого множества мощности больше 1

Распределение Бернулли

Лемма

> **О распределении линейного преобразования дискретной случайной величины**
>
> Пусть $Y = aX + b$, где X - дискретная случайная величина, $a, b \in \mathbb{R}$ $a \neq 0$
> Тогда $PMF_Y(k) = PMF_X(\frac{k-b}{a})$

- Для визуализации обычно используют барчарт

В Python:

```
import scipy.stats as sps
dist = sps.rv_discrete(values=([5, 6, 9], [0.1, 0.2, 0.7]))
dist.pmf(6) #0.2
```

- работает, только если нет повторяющихся значений. Для повторяющихся нужно использовать собственную реализацию:

```
import pandas as pd, scipy.stats as sps
def rv_discrete_pandas(x, p):
    df = pd.DataFrame({'x': x, 'p': p})
    dfg = df.groupby('x', as_index=False).agg({'p': 'sum'})
    return sps.rv_discrete(values=(dfg['x'], dfg['p']))
```

CDF

Определение

> **CDF или кумулятивная функция распределения (cumulative distribution function)**
>
> $$CDF_X(k) = P(X \leq k)$$

- накопленная вероятность всевозможных значений из носителя, не превосходящих заданного числа

Лемма

> **Связь CDF и PMF для дискретного распределения**
>
> $$CDF_X(a) = \sum_{k \le a} PMF_X(k)$$

Лемма

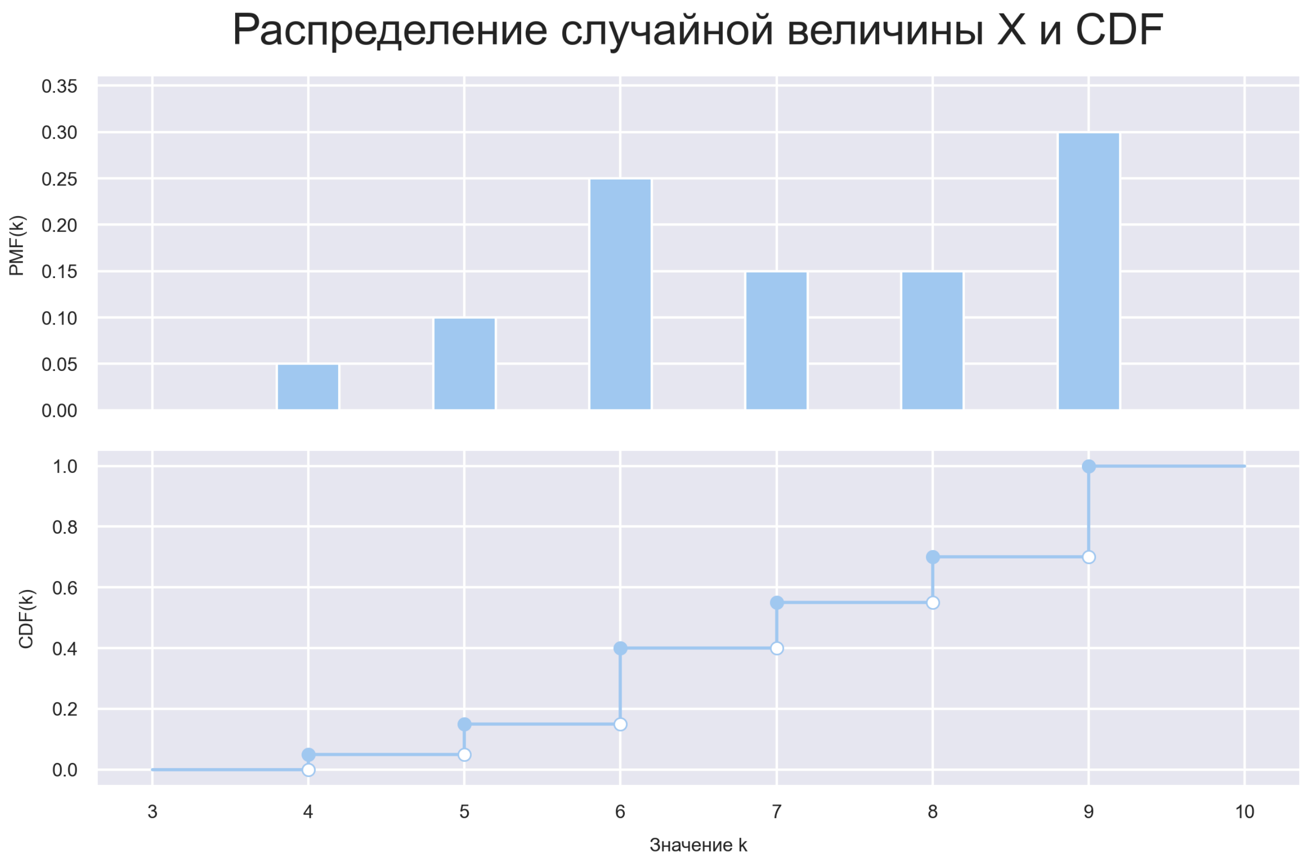
> **Свойства CDF дискретного распределения**
>
> - $P(a \le X \le b) = CDF_X(b) - CDF_X(a_{prev})$
> - $P(X < a) = CDF_X(a_{prev}) = CDF_X(a - \varepsilon)\ \varepsilon \in (0; a - a_{prev})$
> - $P(X > a) = 1 - CDF_X(a)$
> - $P(X \ge a) = 1 - CDF_X(a_{prev})$
> - $CDF(x) = 0 \quad \forall x < min(supp(X)$
> - $CDF(x) = 1 \quad \forall x \ge max(supp(X)$

где $a_{prev} = max\{x \in supp(X) | x < a\}$

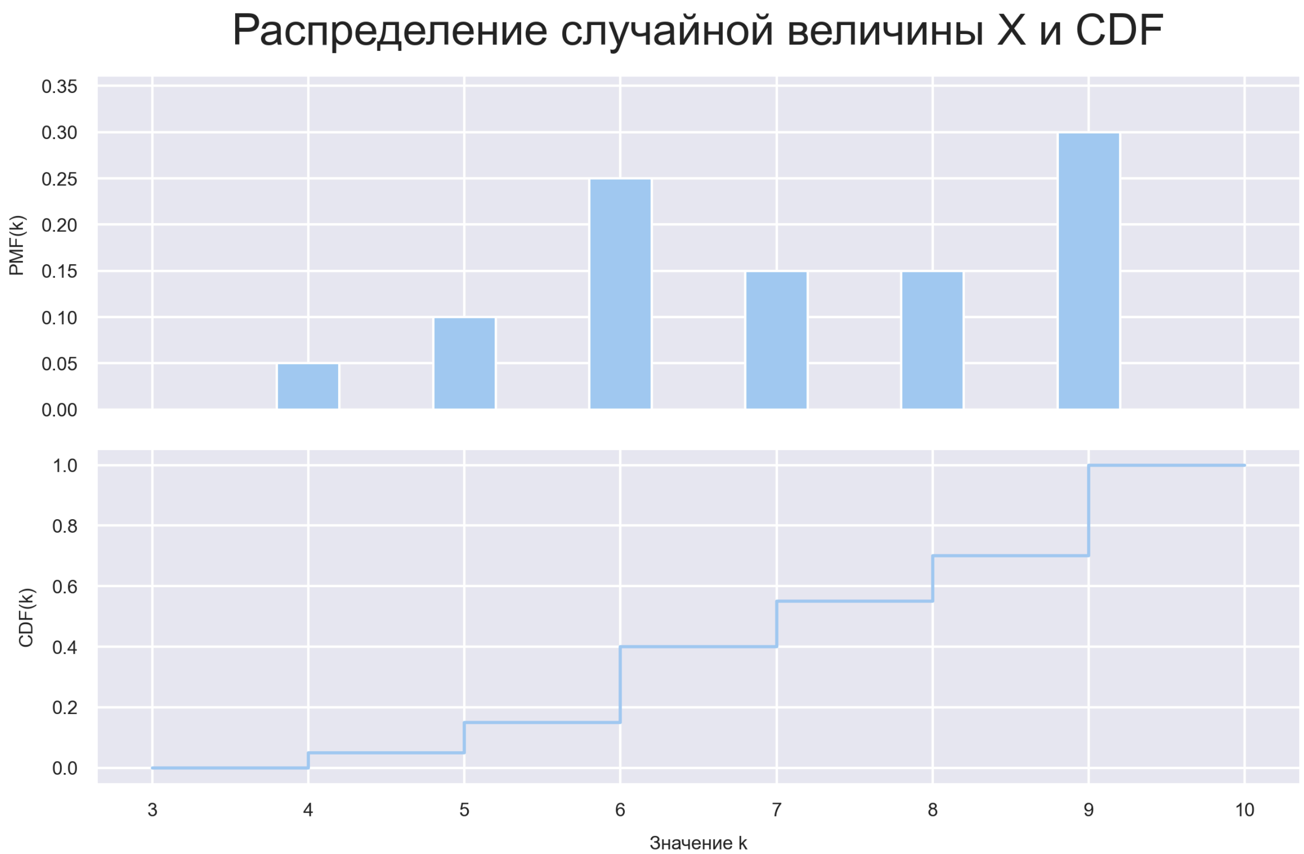
Важно

- по CDF можно однозначно восстановить PMF

- Для визуализации используют линейный график



На практике, выколотыми точками пренебрегают



В Python:

```python
import scipy.stats as sps
dist = sps.rv_discrete(values=([5, 6, 9], [0.1, 0.2, 0.7]))
```

```
dist.cdf(6) #0.3
```

PPF

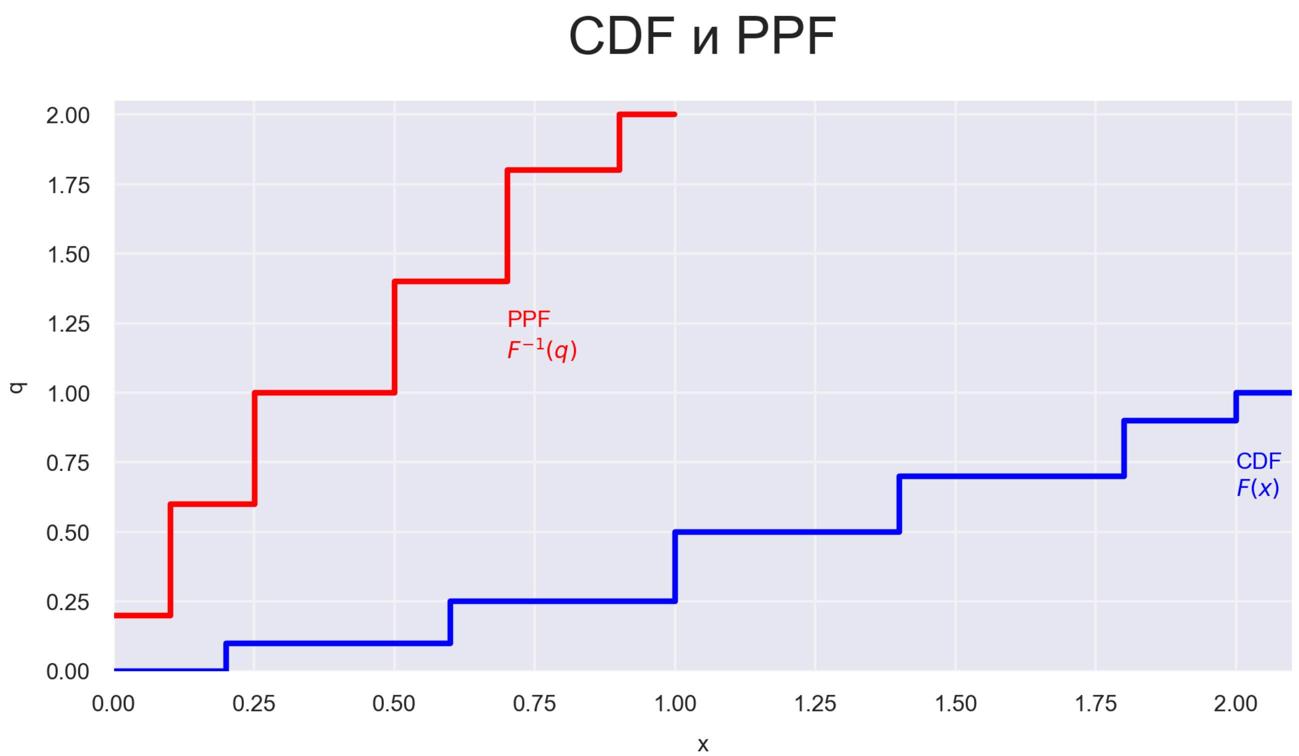
Определение

> **PPF или обратная функция распределения (percentage points function)**
>
> $$PPF_X(q) = a, \text{если} \begin{cases} P(X \le a) \ge q \\ P(X < a) < q \end{cases}$$
>
> или: a - минимальное число, при котором $CDF_X(a) \ge q$

PPF часто используется для нахождения **квантилей** (percentiles). Например, `ppf(0.95)` — 95-й процентиль.

- Для визуализации используется линейный график

Выколотыми точками на практике пренебрегают
Выглядит как отзеркаленный относительно y=x график CDF:



В Python:

```
import scipy.stats as sps
dist = sps.rv_discrete(values=([5, 6, 9], [0.1, 0.2, 0.7]))
dist.ppf(0.5) #9
```

Flashcards

tags: #flashcardsSTAT

Что такое Дискретное вероятностное пространство?
%
— упорядоченная тройка:

- Ω - пространство элементарных исходов
- $\mathcal{F} = 2^{\Omega}$ - множество всех допустимых событий
- $P : \mathcal{F} \to [0; 1]$ - вероятность $P(A) = \sum_{\omega \in A} P(\omega)$

Сформулируйте лемму: Свойства вероятности в дискретном вероятностном пространстве
%

1. $P(A) \geq 0$ для любого события A
2. $P(\Omega) = 1$
3. $P(A_1 \sqcup A_2 \sqcup \ldots \sqcup A_m) = P(A_1) + \ldots + P(A_m)$
4. $P(A_1 \sqcup A_2 \sqcup \ldots) = P(A_1) + P(A_2) + \ldots$ - для счетного числа событий

PMF

Что такое PMF или функция вероятности в точке (probability mass function)?
%

$$PMF_X(k) = P(X = k)$$

Сформулируйте лемму: Свойства PMF в дискретном пространстве
%

$$\begin{gathered} \mathrm{PMF}_X(k) \geq 0 \quad \text{для всех } k \in \mathbb{R}; \\ \sum_{x \in \mathrm{supp}(X)} \mathrm{PMF}_X(x) = 1. \end{gathered}$$

Что такое Дискретное распределение?
%
— распределение, заданное с помощью PMF, удовлетворяющей вышеуказанным свойствам

Что такое Параметр распределения?
%
— параметр PMF, такой, что свойства PMF выполнены при любом значении параметра из некоторого множества мощности больше 1

Сформулируйте лемму: О распределении линейного преобразования дискретной случайной величины
%
Пусть $Y = aX + b$, где X - дискретная случайная величина, $a, b \in \mathbb{R}$ $a \neq 0$
Тогда $PMF_Y(k) = PMF_X(\frac{k-b}{a})$

CDF

Что такое CDF или кумулятивная функция распределения (cumulative distribution function)?
%

$$CDF_X(k) = P(X \leq k)$$

- накопленная вероятность всевозможных значений из носителя, не превосходящих заданного числа

Сформулируйте лемму: Связь CDF и PMF для дискретного распределения
%

$$CDF_X(a) = \sum_{k \leq a} PMF_X(k)$$

Сформулируйте лемму: Свойства CDF дискретного распределения
%

- $P(a \leq X \leq b) = CDF_X(b) - CDF_X(a_{prev})$
- $P(X < a) = CDF_X(a_{prev}) = CDF_X(a - \varepsilon)\ \varepsilon \in (0; a - a_{prev})$
- $P(X > a) = 1 - CDF_X(a)$
- $P(X \geq a) = 1 - CDF_X(a_{prev})$
- $CDF(x) = 0 \quad \forall x < min(supp(X)$
- $CDF(x) = 1 \quad \forall x \geq max(supp(X)$

PPF

Что такое PPF или обратная функция распределения (percentage points function)?
%

$$PPF_X(q) = a, \text{если } \begin{cases} P(X \leq a) \geq q \\ P(X < a) < q \end{cases}$$

или: a - минимальное число, при котором $CDF_X(a) \geq q$