ФКН

Департамент больших данных и информационного поиска

Москва 2025

Теория вероятностей и математическая статистика

Машинное обучение в цифровом продукте

Полякова И.Ю.

О курсе:

- Теория вероятностей и математическая статистика

- И… основы машинного обучения
- И… капелька эконометрики
- И… щепотка анализа данных

План курса (примерный):

1. Введение в теорию вероятностей. Основные понятия
2. Случайные величины и их характеристики.
3. Важнейшие распределения вероятностей
4. Предельные теоремы теории вероятностей
5. Проверка статистических гипотез
6. Байесовский подход к вероятности
7. Линейный модели
8. Теория оценки параметров. Метод максимального правдоподобия (ММП)
9. Смеси распределений и латентные переменные, EM-алгоритм
10. Вероятностные модели в ML
11. Оценка качества и интерпретируемости ML-моделей
12. Оценка и калибровка вероятностей в классификации
13. Байесовские методы выбора и оценки моделей
14. Верификация моделей на практике: стабильность, деградация и A/B-тестирование

Важное:

Формула оценивания:

Микроконтроли * **0.2** + Домашняя контрольная работа 1 * **0.25** +

Домашняя контрольная работа 2 * **0.25** +

Экзамен * **0.3**

Микроконтроли	Активность на семинарах, проявляющаяся в разных формах: решение у доски, выполнение семинарских заданий, подготовка докладов…
ДКР1/ДКР2	«Большие» задания, выполняющиеся в группах в конце 1го и 2го модуля соответственно
Экзамен	Устный, по билетам (не страшный) в конце 2го модуля

Общий чат в телеграм, **подпишись**!

Обо мне:

ВШЭ «Экономика»
Майнор ИАД

- Была научным сотрудником в лабораториях ВШЭ
- Преподавала курсы «Микроэкономика-2» и «Машинное обучение в экономике и финансах»

- Работаю Data Scientist в ПСБ
- Преподаю на ФКН

Контакты:

polyakova.iryu@gmail.com
Telegram: @irra_po
+79194599656

Небольшой тест (не на оценку!)

ФКН

Департамент больших данных и информационного поиска

Москва 2025

Лекция 1
Введение в теорию вероятностей. Основные понятия

Машинное обучение в цифровом продукте

Полякова И.Ю.

Теория вероятностей – математическая наука, изучающая закономерности, присущие массовым случайным явлениям

Предмет изучения: математические модели случайных явлений

Случайные события

Эксперимент: 

- **Случайные события** – любые исходы опыта, которые могут произойти или не произойти (обозначаются заглавными буквами латинского алфавита: A, B, C…)

\- *Выпадение 5 очков, выпадение четного числа очков, выпадение целого числа очков, выпадение не менее 4-х очков…*

- **Элементарные события** – неразложимые и взаимоисключающие исходы w_1, w_2, w_3 … в этом опыте

\- *Выпадение грани «1», выпадение грани «2»…*

- **Пространство элементарных событий (исходов)** – множество всех элементарных событий (обозначается через Ω)

- **Достоверное событие** – событие, которое обязательно наступит в результате данного опыта (Ω)
- **Невозможное событие** – событие, которое заведомо не произойдет в результате проведения опыта (∅)

- **Несовместные события** – появление одного из событий исключает появление другого в одном и том же опыте
- События $A_1, A_2, \ldots, A_n$ – **попарно-несовместны**, если любые два из них несовместны
- Несколько событий образуют **полную группу**, если они попарно несовместные и в результате каждого опыта происходит одно и только одно из них
- **Равновозможные события** – такие события, где ни одно не является объективно более возможным, чем другие

Операции над событиями:

- **Сумма** (A+B или $A \cup B$)

\- Содержит элементы, принадлежащие хотя бы A или B

- **Произведение** (AB или $A \cap B$)

\- Содержит элементы общие для A и B

- **Разность** (A-B или A\B)

\- Содержит элементы события A, не принадлежащие событию B

- **Противоположное событие** ($\bar{A} = \Omega$\A)

Понятие вероятности

Классическое определение вероятности

1. **Конечное число исходов** $\{w_1, w_2, w_3, \dots w_n\}$;
2. **Исходы попарно несовместны**;
3. **Исходы равновозможны**

$$P(w_i) \coloneqq \frac{1}{n}$$

Классическое определение вероятности

Вероятностью события A называется отношение числа **m** случаев , благоприятствующих этому событию, к общему **n** числу случаев, т.е.

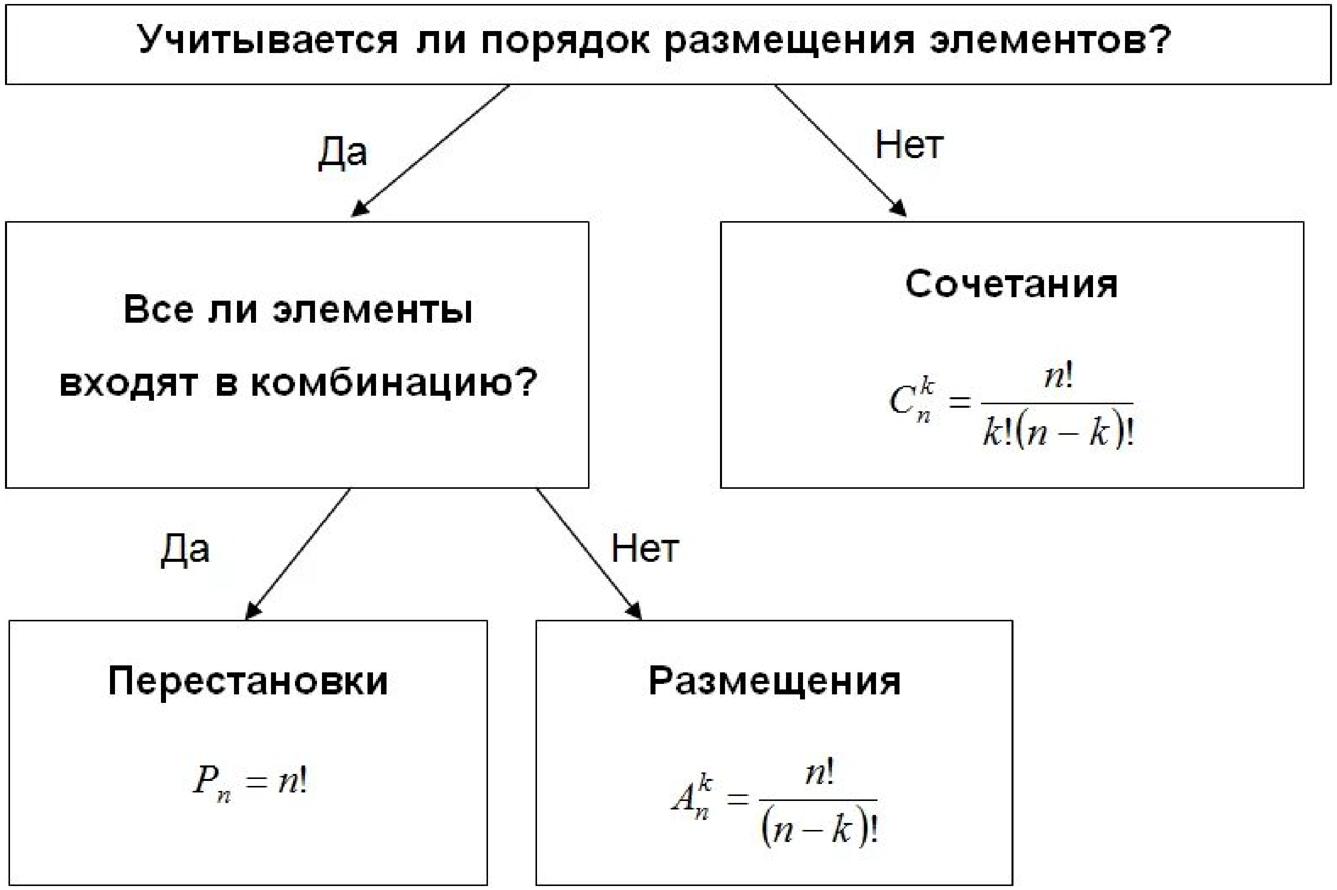
$$\boldsymbol{p = P(A) = \frac{m}{n}}$$

Свойства:

1. $\boldsymbol{P(A) \in [0, 1]}$
2. $\boldsymbol{P(A) = 1 \leftrightarrow A = \Omega}$
3. $\boldsymbol{P(\emptyset) = 0}$
4. $\boldsymbol{P(\overline{A}) = 1 - P(A)}$
5. $\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

Про 5. в случае более двух событий можно посмотреть здесь: Теория вероятностей, Райгородский А.М., Лекция 01, 03.09.20 (формула включений и исключений)

Основные формулы комбинаторики



Для практики

1. В группе 14 девочек и 6 мальчиков. Какова вероятность для работы в паре выбрать двух студентов одного пола?

2. В урне находятся 12 белых и 8 черных шаров. Найти вероятность того, что среди наугад вытянутых 5ти шаров 3 будут черными

Для практики

В группе 14 девочек и 6 мальчиков. Какова вероятность для работы в паре выбрать двух студентов одного пола

$$P(A) = \frac{C_{14}^{2} + C_{6}^{2}}{C_{20}^{2}} = \frac{91 + 15}{190} = \frac{106}{190}$$

Для практики

В урне находятся 12 белых и 8 черных шаров. Найти вероятность того, что среди наугад вытянутых 5ти шаров 3 будут черными

$$\frac{C_8^3 \cdot C_{12}^2}{C_{20}^5} = \frac{56 \times 66}{15504} \approx 0{,}24$$

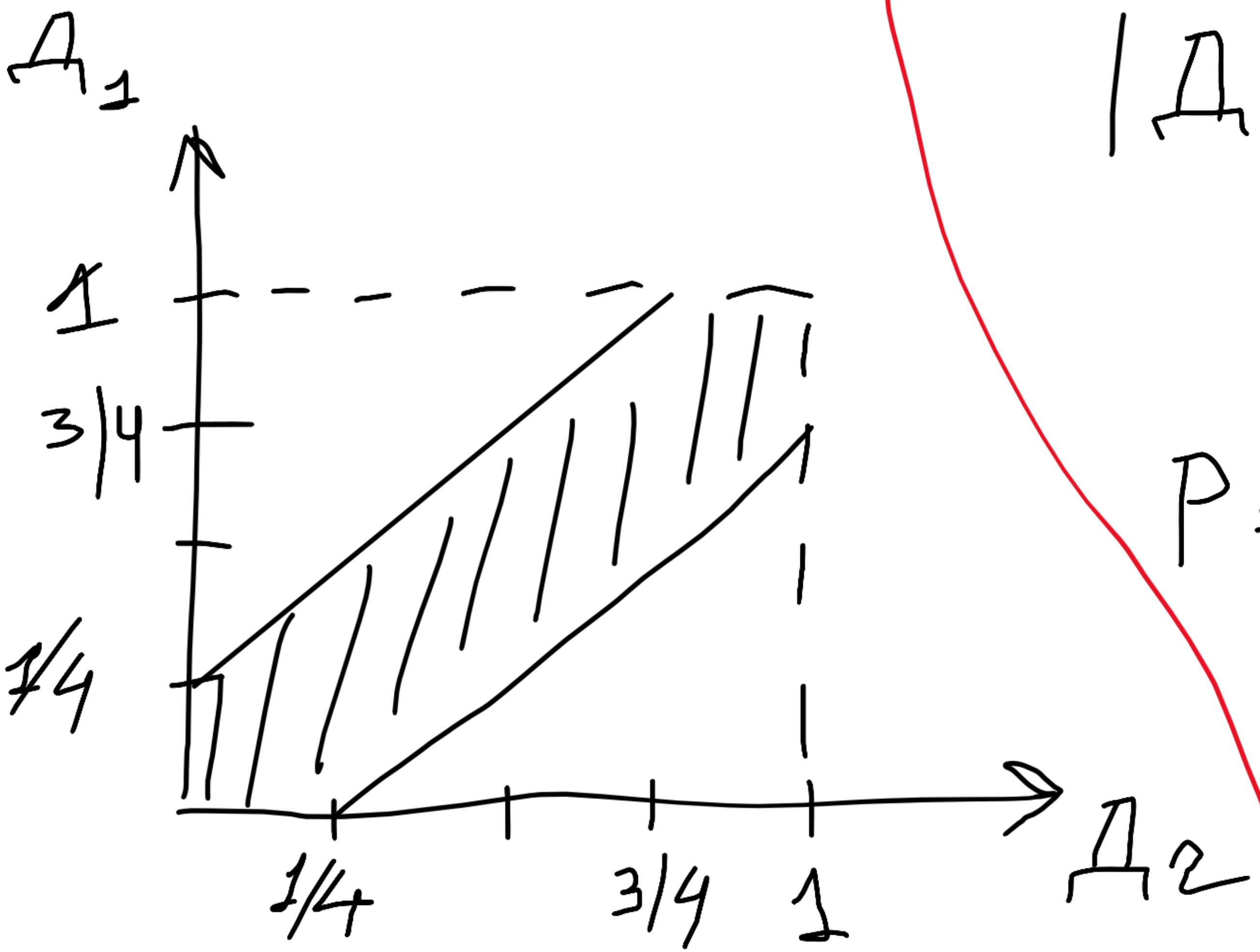
Для практики

1 задание 1 балл

Среди 100 человек, подавших заявку на стажировку, случайным образом выбираются 60. Далее они случайным образом распределяются на 3 команды по 20 человек. Петя и Вася подали заявку на стажировку. Какова вероятность, что они попадут в одну команду?

Источник: Т-Банк, отбор в «Академию аналитиков», 2025

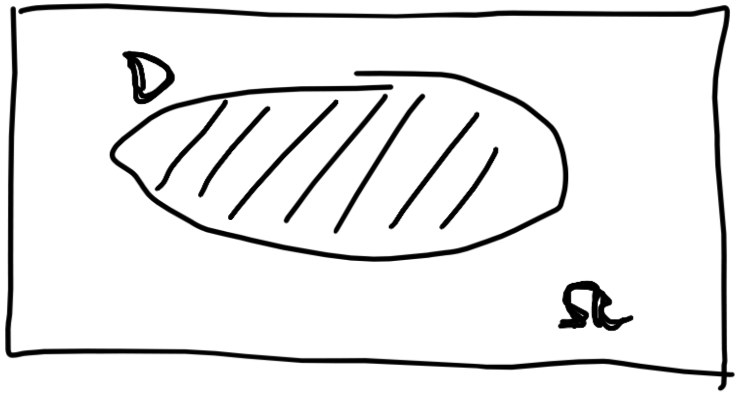
Пусть два друга договорились встретиться с 11 до 12 на остановке.
При этом: каждый из них приходит в случайный момент времени, ждет 15 минут.
Если второй человек не пришел в этот промежуток – встреча считается не состоявшейся. Какова вероятность, что встреча состоится?



$$|Д_1 - Д_2| \le \frac{1}{4}$$

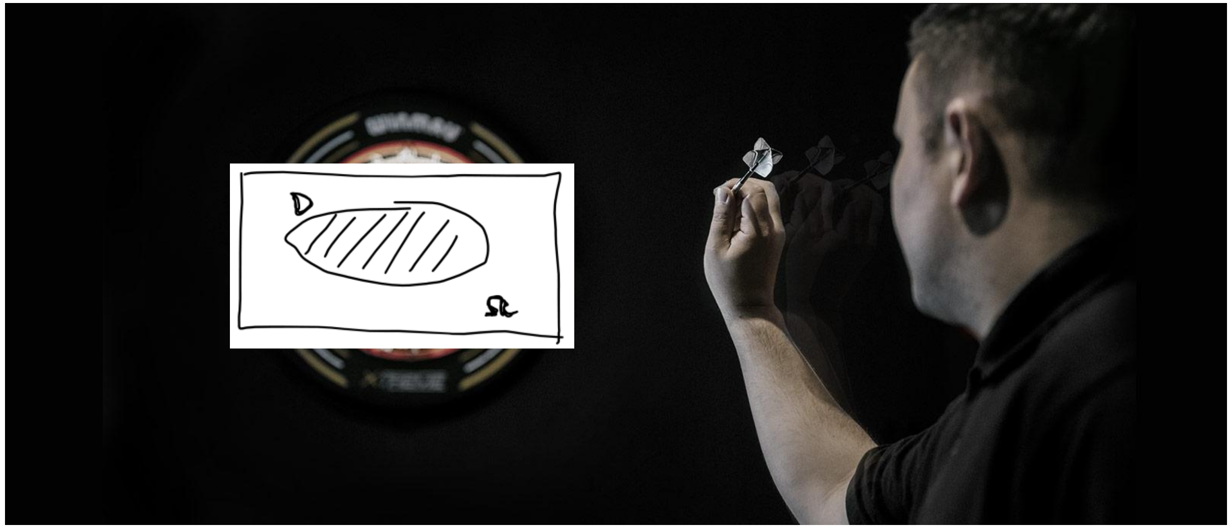
$$P = \frac{\dfrac{1 - \left(\frac{3}{4}\right)^2 \times 2}{2}}{1} = \frac{7}{16} \approx 0{,}44$$

Геометрическое определение вероятности



$$P(A) = \frac{S_D}{S_\Omega}$$

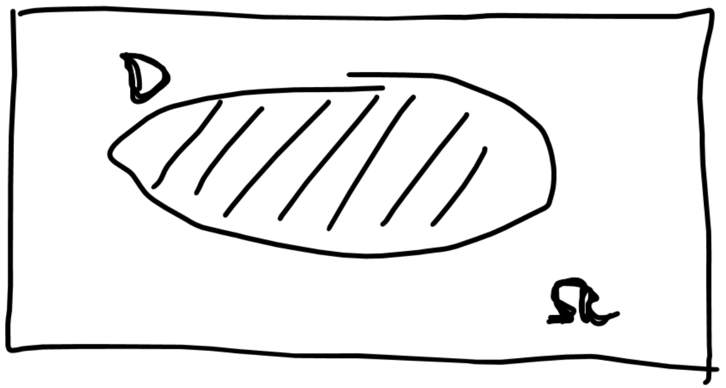
Геометрическое определение вероятности



$$P(A) = \frac{S_D}{S_\Omega}$$

Геометрическое определение вероятности

Пересчитать элементарные события нельзя!



Свойства:

1. $P(A) \in [0, 1]$
2. $P(A) = 1 \leftarrow A = \Omega$
3. $P(\overline{A}) = 1 - P(A)$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$P(A) = \frac{S_D}{S_\Omega}$$

?

Статистическое определение вероятности

Есть эксперимент, который мы можем повторить неограниченное количество раз.

Статистической вероятностью события А будем называть число, около которого колеблется относительная частота события А при достаточно большом количестве испытаний.

Свойства:

1. $\boldsymbol{P(A) \in [0, 1]}$
2. $\boldsymbol{P(A) = 1 \leftarrow A = \Omega}$
3. $\boldsymbol{P(\emptyset) = 0}$
4. $\boldsymbol{P(\overline{A}) = 1 - P(A)}$
5. $\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

$$\boldsymbol{P(A) \approx P^*(A) = \frac{n_A}{n}}$$

$$\boldsymbol{p = \lim_{n \to \infty} \frac{n_A}{n}}$$

Аксиоматическое определение вероятности (обобщение предыдущих)

- Ввел А.Н. Колмогоров

- ***А1. Аксиома неотрицательности***: вероятность любого события А не отрицательна

$$P(A) \geq 0$$

- ***А2. Аксиома нормированности***: вероятность достоверного события равна единице

$$P(\Omega) = 1$$

- ***А3. Аксиома аддитивности***: вероятность суммы несовместных событий равна сумме вероятности этих событий

$$P\left(\sum_k A_k\right) = \sum_k P(A_k)$$

Аксиоматическое определение вероятности (обобщение предыдущих)

- **С1.** Вероятность невозможного события равна 0

$$\boldsymbol{P(\emptyset) = 0}$$

- **С2.** Сумма вероятностей противоположных событий равна единице

$$\boldsymbol{P\left(\overline{A}\right) + P(A) = 1}$$

- **С3.** Вероятность любого события не превосходит единицы

$$\boldsymbol{P(A) \leq 1}$$

- **С4.** Если событие А влечет за собой событие В, то:

$$\boldsymbol{P(A) \leq P(B)}$$

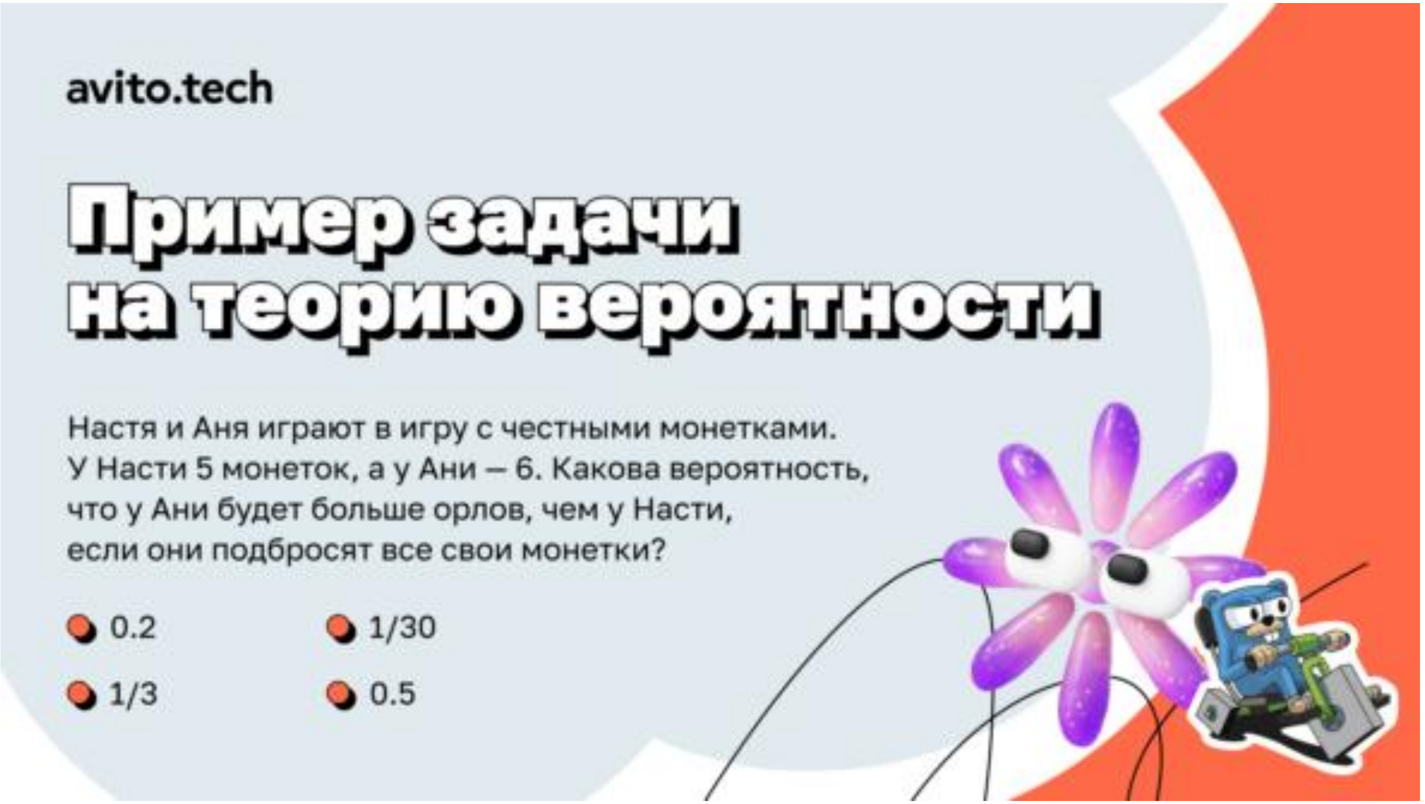
- **С5.** Если события A_1,…, A_n образуют полную группу несовместных событий, то**:**

$$\boldsymbol{\sum_{i=1}^{n} P(\,A_i) = 1}$$

Дополнительно:

- Теория вероятностей, Райгородский А.М., Лекция 01, 03.09.20
- Д. Письменный, Конспект лекций по теории вероятностей, математической статистике и случайным процессам, 2004 (1 глава)
- Вероятности вероятностей: #2. Нулевая вероятность не значит «невозможно» [3Blue1Brown]

Дополнительно для практики



Источник: https://vc.ru/hr/1770701-kak-popast-na-stazhirovku-dlya-analitikov-v-avito-gaid-po-etapam-otbora

ВЫСШАЯ • ШКОЛА • ЭКОНОМИКИ •