Упорядоченные выборки, с повторениями и без

Владимир Подольский

Факультет компьютерных наук, Высшая Школа Экономики

Упорядоченные выборки

Количество слов

Язык множеств для последовательностей

Автомобильные номера

Выбор с ограничениями

Перестановки

Задача

Сколько существует различных паролей, состоящих из пяти строчных латинских букв (всего латинских букв 26)?

Задача

Сколько существует различных паролей, состоящих из пяти строчных латинских букв (всего латинских букв 26)?

 Оказывается, что для решения этой задачи достаточно знать правило произведения

Задача

Сколько существует различных паролей, состоящих из пяти строчных латинских букв (всего латинских букв 26)?

- Оказывается, что для решения этой задачи достаточно знать правило произведения
- Но мы должны его применять последовательно шаг за шагом

• Начнем с подсчета паролей из одной буквы

- Начнем с подсчета паролей из одной буквы
- Ясно, что их 26

- Начнем с подсчета паролей из одной буквы
- Ясно, что их 26
- А что с двухбуквенными паролями?

- Начнем с подсчета паролей из одной буквы
- Ясно, что их 26
- А что с двухбуквенными паролями?
- Теперь мы можем выбрать каждую из букв 26 способами

26 26

* *

- Начнем с подсчета паролей из одной буквы
- Ясно, что их 26
- А что с двухбуквенными паролями?
- Теперь мы можем выбрать каждую из букв 26 способами
- Применяем правило произведения, получаем 676 вариантов

$$26 \times 26 = 676$$

* *

• Перейдем к случаю трехбуквенных паролей

- Перейдем к случаю трехбуквенных паролей
- Мы уже знаем, что первые две буквы можно выбрать 676 способами



- Перейдем к случаю трехбуквенных паролей
- Мы уже знаем, что первые две буквы можно выбрать 676 способами
- Третью букву можно выбрать вновь 26 способами

- Перейдем к случаю трехбуквенных паролей
- Мы уже знаем, что первые две буквы можно выбрать 676 способами
- Третью букву можно выбрать вновь 26 способами
- И мы снова применяем правило произведения!

- Перейдем к случаю трехбуквенных паролей
- Мы уже знаем, что первые две буквы можно выбрать 676 способами
- Третью букву можно выбрать вновь 26 способами
- И мы снова применяем правило произведения!
- Получаем 17 576 вариантов

$$676 \times 26 = 17576$$
 $* * *$

• Точно также мы можем продолжать и дальше

• Точно также мы можем продолжать и дальше

• Точно также мы можем продолжать и дальше

- Точно также мы можем продолжать и дальше
- В итоге получаем 11 881 376 способов составить пароль из 5 букв

$$26 \times 26 \times 26 \times 26 \times 26 = 11881376$$
* * * * * *

Задача

Пусть у нас есть множество из n символов. Сколько различных последовательностей длины k можно составить из этих символов?

Задача

Пусть у нас есть множество из n символов. Сколько различных последовательностей длины k можно составить из этих символов?

Такие последовательности обычно называются словами или упорядоченными выборками с повторениями

• Мы можем применить те же рассуждения

Задача

Пусть у нас есть множество из n символов. Сколько различных последовательностей длины k можно составить из этих символов?

- Мы можем применить те же рассуждения
- Есть n способов выбрать первую букву

Задача

Пусть у нас есть множество из n символов. Сколько различных последовательностей длины k можно составить из этих символов?

- Мы можем применить те же рассуждения
- Есть n способов выбрать первую букву
- Выбор каждой следующей буквы увеличивает количество последовательностей в n раз

Задача

Пусть у нас есть множество из n символов. Сколько различных последовательностей длины k можно составить из этих символов?

- Мы можем применить те же рассуждения
- Есть n способов выбрать первую букву
- Выбор каждой следующей буквы увеличивает количество последовательностей в n раз
- Количество последовательностей равно числу n, умноженному само на себя k раз, то есть n^k

Упорядоченные выборки

Количество слов

Язык множеств для последовательностей

Автомобильные номера

Выбор с ограничениями

Перестановки

 Для последовательностей есть обозначения на языке множеств

- Для последовательностей есть обозначения на языке множеств
- Пусть нам даны два множества A и B

- Для последовательностей есть обозначения на языке множеств
- Пусть нам даны два множества A и B
- Через $A \times B$ мы обозначаем множество всех пар (a,b), где $a \in A$ и $b \in B$

- Для последовательностей есть обозначения на языке множеств
- Пусть нам даны два множества A и B
- Через $A \times B$ мы обозначаем множество всех пар (a,b), где $a \in A$ и $b \in B$
- Множество $A \times B$ называется декартовым произведением множеств A и B

• Если A и B конечны, то количество элементов в $A \times B$ равно $|A| \cdot |B|$

- Если A и B конечны, то количество элементов в $A \times B$ равно $|A| \cdot |B|$
- Другими словами, $|A \times B| = |A| \cdot |B|$

- Если A и B конечны, то количество элементов в $A \times B$ равно $|A| \cdot |B|$
- Другими словами, $|A \times B| = |A| \cdot |B|$
- Это просто переформулировка правила произведения на языке множеств

- В общем виде, пусть у нас есть множества A_1,A_2,\ldots,A_k

- В общем виде, пусть у нас есть множества A_1, A_2, \dots, A_k
- Через $A_1 imes A_2 imes ... imes A_k$ мы обозначаем множество всех последовательностей (a_1,a_2,\ldots,a_k) , где $a_1 \in A_1$, $a_2 \in A_2$ и так далее

- В общем виде, пусть у нас есть множества A_1,A_2,\dots,A_k
- Через $A_1 imes A_2 imes ... imes A_k$ мы обозначаем множество всех последовательностей (a_1,a_2,\ldots,a_k) , где $a_1 \in A_1$, $a_2 \in A_2$ и так далее
- Множество $A_1 \times A_2 \times ... \times A_k$ называется декартовым произведением множеств $A_1, A_2, ..., A_k$

• В случае $A_1 = A_2 = ... = A_k$ удобно сокращать обозначение до A^k

- В случае $A_1=A_2=\ldots=A_k$ удобно сокращать обозначение до A^k
- Другими словами, множество последовательностей длины k, в которых каждый символ выбирается из множества A, обозначается A^k

• Мы видели в прошлом видео, что для конечного множества A

$$|A^k| = |A|^k$$

• Мы видели в прошлом видео, что для конечного множества ${\cal A}$

$$|A^k| = |A|^k$$

• На самом деле, полностью аналогично можно показать, что для конечных A_1, A_2, \dots, A_k

$$|A_1 \times A_2 \times \ldots \times A_k| = |A_1| \cdot |A_2| \cdot \ldots \cdot |A_k|$$

• Мы видели в прошлом видео, что для конечного множества A

$$|A^k| = |A|^k$$

• На самом деле, полностью аналогично можно показать, что для конечных A_1, A_2, \dots, A_k

$$|A_1 \times A_2 \times \ldots \times A_k| = |A_1| \cdot |A_2| \cdot \ldots \cdot |A_k|$$

• Мы увидим пример в следующем видео

• Слова встречаются в анализе данных очень часто

- Слова встречаются в анализе данных очень часто
- Пусть в наших данных есть признаки «марка машины», «модель телефона», «профессия»

- Слова встречаются в анализе данных очень часто
- Пусть в наших данных есть признаки «марка машины», «модель телефона», «профессия»
- Тогда каждый человек кодируется в наших данных словом

- Слова встречаются в анализе данных очень часто
- Пусть в наших данных есть признаки «марка машины», «модель телефона», «профессия»
- Тогда каждый человек кодируется в наших данных словом
- Например, (УАЗ Хантер, Nokia 3310, егерь) может быть объектом в наших данных

Упорядоченные выборки

Количество слов

Язык множеств для последовательностей

Автомобильные номера

Выбор с ограничениями

Перестановки

- C 065 MK 78-

wikimedia.org

• Теперь мы готовы обсудить обещанный пример

· C 065 MK 78.

- Теперь мы готовы обсудить обещанный пример
- Автомобильные номера: 3 цифры, 3 буквы; 78 код региона

· C 065 MK 78-

- Теперь мы готовы обсудить обещанный пример
- Автомобильные номера: 3 цифры, 3 буквы; 78 код региона
- Возможности: 10 вариантов цифр, 12 вариантов букв (используются только те буквы, которые похожи на аналогичные латинские)

· C 065 MK 78-

- Теперь мы готовы обсудить обещанный пример
- Автомобильные номера: 3 цифры, 3 буквы; 78 код региона
- Возможности: 10 вариантов цифр, 12 вариантов букв (используются только те буквы, которые похожи на аналогичные латинские)
- Сколько возможных номеров есть для каждого региона?

- C 065 MK 78-

wikimedia.org

• Каждую цифру можно выбрать 10 способами

· C 065 MK 78.

- Каждую цифру можно выбрать 10 способами
- Так что последовательность цифр можно выбрать $10 \times 10 \times 10 = 1000$ способами

· C 065 MK 78-

- Каждую цифру можно выбрать 10 способами
- Так что последовательность цифр можно выбрать $10 \times 10 \times 10 = 1000$ способами
- Каждую букву можно выбрать 12 способами

· C 065 MK 78-

- Каждую цифру можно выбрать 10 способами
- Так что последовательность цифр можно выбрать $10 \times 10 \times 10 = 1000$ способами
- Каждую букву можно выбрать 12 способами
- Так что последовательность цифр можно выбрать $12 \times 12 \times 12 = 1728$ способами

· C 065 MK 78.

- Каждую цифру можно выбрать 10 способами
- Так что последовательность цифр можно выбрать $10 \times 10 \times 10 = 1000$ способами
- Каждую букву можно выбрать 12 способами
- Так что последовательность цифр можно выбрать $12 \times 12 \times 12 = 1728$ способами
- Всего получается 1728 000 номеров для региона

· C 065 MK 78.

wikimedia.org

• Получается 1728 000 номеров для региона

· C 065 MK 78.

- Получается 1 728 000 номеров для региона
- Это оценка сверху: не все комбинации букв и цифр используются

- C 065 MK 78-

- Получается 1 728 000 номеров для региона
- Это оценка сверху: не все комбинации букв и цифр используются
- Достаточно ли этого?

· C 065 MK 78.

- Получается 1728 000 номеров для региона
- Это оценка сверху: не все комбинации букв и цифр используются
- Достаточно ли этого?
- Не всегда: например, в Москве зарегистрировано около 5 600 000 автомобилей (на 2016 год)

- C 065 MK 78-

wikimedia.org

• Как решается эта проблема?

· C 065 MK 78.

- Как решается эта проблема?
- Для одного региона вводится несколько кодов региона

· C 065 MK 78.

- Как решается эта проблема?
- Для одного региона вводится несколько кодов региона
- В Москве их девять

· C 065 MK 78.

- Как решается эта проблема?
- Для одного региона вводится несколько кодов региона
- В Москве их девять
- Потребовались трехзначные коды региона

Упорядоченные выборки

Количество слов

Язык множеств для последовательностей

Автомобильные номера

Выбор с ограничениями

Перестановки

Выбор с ограничениями

 Мы видели как с помощью правила произведения можно посчитать количество слов заданной длины в заданном алфавите

Выбор с ограничениями

- Мы видели как с помощью правила произведения можно посчитать количество слов заданной длины в заданном алфавите
- Но правило произведения позволяет подсчитывать и другие объекты

Задача

Сколько существует целых чисел от 0 до 9999, в которых есть ровно одна цифра 7?

Задача

Сколько существует целых чисел от 0 до 9999, в которых есть ровно одна цифра 7?

 Числа от 0 до 9999 — это последовательности цифр длины 4

Задача

Сколько существует целых чисел от 0 до 9999, в которых есть ровно одна цифра 7?

- Числа от 0 до 9999 это последовательности цифр длины 4
- Трехзначные числа соответствуют последовательностям, начинающимся с нуля 0

 Мы можем поместить единственную цифру 7 на любую из 4 позиций

- Мы можем поместить единственную цифру 7 на любую из 4 позиций
- Это дает 4 случая; если мы посчитаем количество последовательностей в каждом из 4 случаев, мы сможем найти ответ по правилу суммы

- Мы можем поместить единственную цифру 7 на любую из 4 позиций
- Это дает 4 случая; если мы посчитаем количество последовательностей в каждом из 4 случаев, мы сможем найти ответ по правилу суммы
- Рассмотрим один из случаев

- Мы можем поместить единственную цифру 7 на любую из 4 позиций
- Это дает 4 случая; если мы посчитаем количество последовательностей в каждом из 4 случаев, мы сможем найти ответ по правилу суммы
- Рассмотрим один из случаев
- Каждую из оставшихся цифр можно выбрать 9 способами! (цифра 7 запрещена)

• Значит в этом случае получается $9 \times 9 \times 9 = 729$ чисел

- Значит в этом случае получается $9 \times 9 \times 9 = 729$ чисел
- И во всех остальных случаях тоже!

- Значит в этом случае получается $9 \times 9 \times 9 = 729$ чисел
- И во всех остальных случаях тоже!
- Всего у нас 4 случая, так что всего есть $4 \times 729 = 2916$ чисел меньших 10 000 с ровно одной цифрой 7

- Значит в этом случае получается $9 \times 9 \times 9 = 729$ чисел
- И во всех остальных случаях тоже!
- Всего у нас 4 случая, так что всего есть $4 \times 729 = 2916$ чисел меньших 10 000 с ровно одной цифрой 7
- Это меньше 1/3, но больше 1/4 всех четырехзначных чисел

- Значит в этом случае получается $9 \times 9 \times 9 = 729$ чисел
- И во всех остальных случаях тоже!
- Всего у нас 4 случая, так что всего есть $4 \times 729 = 2916$ чисел меньших 10 000 с ровно одной цифрой 7
- Это меньше 1/3, но больше 1/4 всех четырехзначных чисел
- Оценили вероятность получить число с ровно одной цифрой 7 при выборе числа <10 000 "случайно"

Упорядоченные выборки

Количество слов

Язык множеств для последовательностей

Автомобильные номера

Выбор с ограничениями

• Мы обсудили как подсчитывать количество слов

- Мы обсудили как подсчитывать количество слов
- Теперь мы готовы перейти ко второй стандартной комбинаторной постановке: перестановкам

Задача

Пусть у нас есть алфавит из n символов. Сколько есть различных слов длины k в этом алфавите, в которых никакой символ не повторяется дважды?

• Слова длины k без повторений букв называются k-перестановками

Задача

Пусть у нас есть алфавит из n символов. Сколько есть различных слов длины k в этом алфавите, в которых никакой символ не повторяется дважды?

- Слова длины k без повторений букв называются k-перестановками
- Легко видеть, что если n < k, то k-перестановок нет: нам просто не хватит букв в алфавите

Задача

Пусть у нас есть алфавит из n символов. Сколько есть различных слов длины k в этом алфавите, в которых никакой символ не повторяется дважды?

- Слова длины k без повторений букв называются k-перестановками
- Легко видеть, что если n < k, то k-перестановок нет: нам просто не хватит букв в алфавите
- Так что достаточно решить задачу для случая $k \leq n$

• Применим правило произведения (опять!)

- Применим правило произведения (опять!)
- Первую букву можно выбрать n способами

- Применим правило произведения (опять!)
- Первую букву можно выбрать n способами
- Сколько есть способов выбрать вторую букву?

```
1 \qquad 2 \qquad 3 \qquad \dots \qquad k
* \qquad * \qquad * \qquad \cdots \qquad *
n
```

- Применим правило произведения (опять!)
- Первую букву можно выбрать n способами
- Сколько есть способов выбрать вторую букву?
- Мы можем выбрать ее любой, кроме уже занятой буквы на первой позиции

- Применим правило произведения (опять!)
- Первую букву можно выбрать n способами
- Сколько есть способов выбрать вторую букву?
- Мы можем выбрать ее любой, кроме уже занятой буквы на первой позиции
- Первая буква может быть любой, но в любом случае вторую букву можно выбрать $n{-}1$ способом!

• Так что первую и вторую букву можно выбрать $n \times (n-1)$ способами

- Так что первую и вторую букву можно выбрать $n \times (n-1)$ способами
- Далее, третью букву можно выбрать $n\!-\!2$ способами: доступны все буквы, кроме уже занятых на первых двух позициях

- Так что первую и вторую букву можно выбрать $n \times (n-1)$ способами
- Далее, третью букву можно выбрать $n\!-\!2$ способами: доступны все буквы, кроме уже занятых на первых двух позициях
- И так далее; для каждой следующей буквы вариантов на один меньше

- Так что первую и вторую букву можно выбрать $n \times (n-1)$ способами
- Далее, третью букву можно выбрать $n\!-\!2$ способами: доступны все буквы, кроме уже занятых на первых двух позициях
- И так далее; для каждой следующей буквы вариантов на один меньше
- Для последней буквы останется $n{-}k{+}1$ вариантов

• Всего мы получили $n \times (n-1) \times ... \times (n-k+1)$ k-перестановок

- Всего мы получили $n \times (n-1) \times ... \times (n-k+1)$ k-перестановок
- Удобное обозначение: $n! = 1 \times 2 \times ... \times n$; это число называется факториалом n

- Всего мы получили $n \times (n-1) \times ... \times (n-k+1)$ k-перестановок
- Удобное обозначение: $n! = 1 \times 2 \times ... \times n$; это число называется факториалом n
- В этих обозначениях число k-перестановок на n символах выглядит лучше: $n!/(n\!-\!k)!$

- Всего мы получили $n \times (n-1) \times ... \times (n-k+1)$ k-перестановок
- Удобное обозначение: $n! = 1 \times 2 \times ... \times n$; это число называется факториалом n
- В этих обозначениях число k-перестановок на n символах выглядит лучше: $n!/(n\!-\!k)!$
- А что если $n{-}k{=}0$?

- Всего мы получили $n \times (n-1) \times ... \times (n-k+1)$ k-перестановок
- Удобное обозначение: n! = $1 \times 2 \times ... \times n$; это число называется факториалом n
- В этих обозначениях число k-перестановок на n символах выглядит лучше: $n!/(n\!-\!k)!$
- А что если n-k=0? соглашение: 0!=1

Задача

Сколько есть разных порядков, в которых можно расставить n разных книг на полке?

Задача

Сколько есть разных порядков, в которых можно расставить n разных книг на полке?

• Каждая книга — это буква

Задача

Сколько есть разных порядков, в которых можно расставить n разных книг на полке?

- Каждая книга это буква
- Нам надо посчитать n-перестановки из n букв; это называется просто перестановками

Задача

Сколько есть разных порядков, в которых можно расставить n разных книг на полке?

- Каждая книга это буква
- Нам надо посчитать n-перестановки из n букв; это называется просто перестановками
- По предыдущей задаче их количество равно n!

 Мы обсудили две стандартные постановки в комбинаторике: слова и перестановки

- Мы обсудили две стандартные постановки в комбинаторике: слова и перестановки
- Все рассуждения опирались на правило произведения

- Мы обсудили две стандартные постановки в комбинаторике: слова и перестановки
- Все рассуждения опирались на правило произведения
- Эти постановки помогают во многих случаях

- Мы обсудили две стандартные постановки в комбинаторике: слова и перестановки
- Все рассуждения опирались на правило произведения
- Эти постановки помогают во многих случаях
- Но они не закрывают все наши потребности

- Мы обсудили две стандартные постановки в комбинаторике: слова и перестановки
- Все рассуждения опирались на правило произведения
- Эти постановки помогают во многих случаях
- Но они не закрывают все наши потребности
- В следующем уроке мы увидим еще одну стандартную постановку