

Байесовская оптимизация для вывода демографических историй

Промежуточная презентация

Илья Шешуков

Руководители: Екатерина Носкова (ИТМО),

Вячеслав Боровицкий (СПбГУ)

Демографическая модель популяции

Имея геномы людей, хотим понять как изменялись их популяции. Как менялась численность, когда популяции разделялись, как сильно они мигрировали.

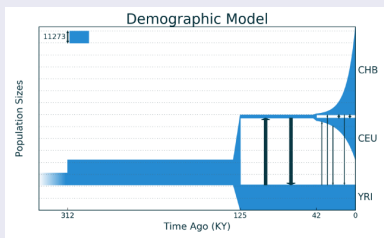


Рис. 1: Популяционная модель человеческой миграции из Африки

Аллель-частотный спектр

Определение (Аллель-частотный спектр)

Аллель-частотный спектр это распределение частоты аллелей в данных локусах в популяции или выборке.

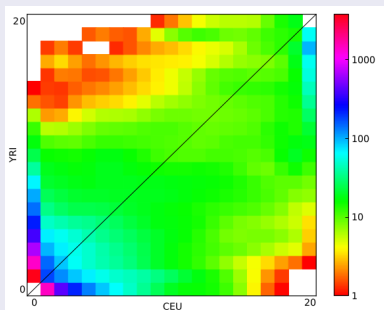


Рис. 2: Хитмэп аллель-частотного спектра двух популяций

Пример

| | <i>SNP1</i> | <i>SNP2</i> | <i>SNP3</i> | <i>SNP4</i> | <i>SNP5</i> | <i>SNP6</i> | <i>SNP7</i> | <i>SNP8</i> |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| Сумма | 1 | 2 | 3 | 1 | 1 | 2 | 5 | 1 |

Спектр: (4 2 1 0 1)

Как это делается сейчас

dadi

<https://bitbucket.org/gutenkunstlab/dadi/>

- Плюсы
 - Она работает
 - Ей пользуются реальные люди
- Минусы
 - Решает дифференциальное уравнение в частных производных, что долго
 - Использует методы локальной оптимизации, что малоэффективно
 - Для работы необходимо руками писать Питон

moments

<https://bitbucket.org/simongravel/moments>

- Плюсы
 - Эффективнее, чем *dadi*, особенно на больших популяциях

GADMA

<https://github.com/ctlab/GADMA>

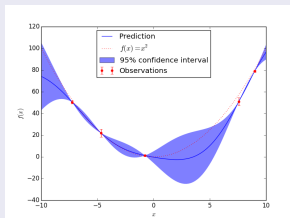
- Основана на *dad*i и moments
- Использует генетический алгоритм для поиска значения параметров демографической модели
- Не требует человеческого вмешательства

Что можно сделать

Давайте попробуем заменить генетический алгоритм на байесовскую оптимизацию

Байесовская оптимизация

- Алгоритм глобальной оптимизации
- Хорошо работает для сложновычислимых функций (например, если нужно решать уравнение в частных производных), т.е. хорошо подходит для задачи
- Можно параллелить
- Менее эвристична, чем генетический алгоритм



Планы

- 1 (В процессе) Заменить в *daDi* алгоритм градиентного спуска на байесовскую оптимизацию.
- 2 Посмотреть станет ли лучше
- 3 (Может быть?) Интегрировать в GADMA

Конец

Спасибо за внимание