

Байесовская оптимизация для вывода демографических историй

Промежуточная презентация

Илья Шешуков

Руководители: Екатерина Носкова (ИТМО), Вячеслав Боровицкий
(СПбГУ)

Введение

Демографическая модель популяции

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

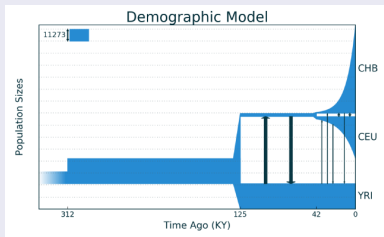


Figure: Out of Africa model

Аллель-частотный спектр

Определение (Аллель-частотный спектр)

Allele frequency spectrum is the distribution of the allele frequencies of a given set of loci (often SNPs) in a population or sample^a.

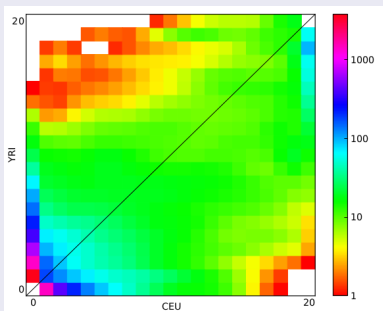


Figure: Хитмэп аллель-частотного спектра двух популяций

^aWikipedia

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
 - Теоретически, для этой задачи можно использовать более эффективный алгоритм
 - Для работы необходимо руками писать Питон

moments

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

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

GADMA

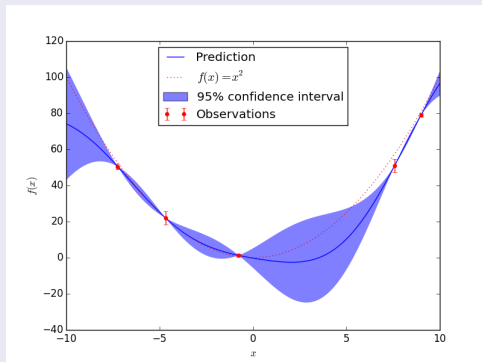
<https://github.com/ctlab/GADMA>

- Основана на *dad*i и moments
- Использует генетический алгоритм
- Не требует человеческого вмешательства

Что можно сделать

Байесовская оптимизация

- Хорошо работает для сложновычислимых функций (например, если нужно решать уравнение в частных производных), т.е. хорошо подходит для задачи
- Можно параллелить



Планы

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

Конец

Спасибо за внимание