

# Исследование биологической модели стационарных сообществ

по работам Михайловой Ксении и Давида Арутюнян

Соколова Диана,  
Лебедюк Ева,  
Василева Анна



## Динамика пространственных моментов

$$\begin{cases} \frac{dN}{dt} = (b - d)N - d' \int_{\mathbb{R}^n} C(\xi) w(\xi) d\xi, \\ \frac{dC(\xi)}{dt} = bm(\xi)N + \int_{\mathbb{R}^n} bm(\xi') C(\xi + \xi') d\xi' - (d + d'w(\xi))C(\xi) - \int_{\mathbb{R}^n} d'w(\xi') T(\xi, \xi') d\xi' \end{cases}$$

- Данная система в точности описывает стохастический процесс и может расширяться бесконечно, то есть до пространственного момента порядка  $k$ , и будет являться системой из  $k+1$  уравнений
- Динамика первого пространственного момента зависит от второго пространственного момента, а второго — от третьего и т.д



## Метод замыкания

Пространственный момент наибольшего порядка выражается через остальные, тем самым уменьшая количество неизвестных

Будем рассматривать следующее трехпараметрическое замыкание:

$$T_{\alpha,\beta,\gamma}(\xi, \xi') = \frac{\alpha C(\xi)C(\xi') + \beta C(\xi)C(\xi' - \xi) + \gamma C(\xi')C(\xi' - \xi) - \beta N^4}{(\alpha + \gamma)N},$$

где  $\alpha, \beta, \gamma \in \mathbb{R}$ .

При условии стационарности динамики первого и второго моментов получаем:

$$\begin{cases} 0 = (b - d)N - \langle C, \bar{w} \rangle, \\ 0 = \bar{m}N + [C * \bar{m}] - (d + \bar{w})C - \frac{1}{(\alpha + \gamma)N} (\alpha C \langle C, \bar{w} \rangle + \beta C[C * \bar{w}] + \gamma [C * \bar{C}w] - \beta N^4) \end{cases}$$

# Метод Неймана поиска оптимальных параметров замыкания

С помощью данного метода можно получить численное решение (если оно существует)

Для начала стоит преобразовать систему, чтобы решение стало исчезать на бесконечности. После некоторых преобразований получим уравнение:

$$Q = \frac{\bar{m} + [\bar{m} * Q] - \bar{w}Q - \frac{\beta}{\alpha + \gamma} Q[\bar{w} * Q] - \frac{\gamma}{\alpha + \gamma} [Q\bar{w} * Q] + \frac{\beta d'}{\alpha + \gamma} N^2}{d + \frac{\alpha(b - d)}{\alpha + \gamma}}$$

где  $Q = \frac{c}{N}$ .

$$Q = \mathcal{A}Q$$

где оператор  $\mathcal{A}$  действует по правилу:

$$\mathcal{A}f = \frac{\bar{m} + [\bar{m} * f] - \bar{w}f - \frac{\beta}{\alpha + \gamma} f[\bar{w} * f] - \frac{\gamma}{\alpha + \gamma} [f\bar{w} * f] + \frac{\beta d'}{\alpha + \gamma} (Nf)^2}{d + \frac{\alpha(b - d)}{\alpha + \gamma}}$$



## Градиентный метод

Встает задача об оптимизации значения второго пространственного момента по  $\alpha, \beta, \gamma$ , полученного с помощью метода Неймана,

$$\min_{\alpha, \beta, \gamma} L = \left\| \sum_{k=1}^m \mathcal{A}_{\alpha, \beta, \gamma} C - C_{sim} \right\|_2^2,$$

где  $C_{sim}$  — значение эталонного второго пространственного момента и  $m$  — число итераций метода Неймана. Шаг проекции градиента:

$$x_{i+1} = P_X(x_i - \lambda_i \nabla L(x_i)),$$

где  $x = (\alpha, \beta, \gamma)$ ,  $X = \{x \in \mathbb{R}^3 : \langle x, (1, 1, 1) \rangle = 1\}$ ,  $P_X(x) = x + \frac{1 - \langle x, (1, 1, 1) \rangle}{3} (1, 1, 1)$ .

$\lambda_i$  — длина шага, которую можно выбрать любую, например, постоянный шаг или наискорейший спуск.



## Результаты

Ниже приведена таблица с входными параметрами  $b, d, d', sm$  и  $sw$  и получившимся оптимальными параметрами замыкания  $\alpha, \beta, \gamma$ :

ID	$b$	$d$	$d'$	$sm$	$sw$	$\alpha$	$\beta$	$\gamma$
1	0.2	0.02	0.02	0.96	0.09	2.8077	-0.57713	-1.23058
2	1	0.2	0.75	0.96	0.09	3.05231	-1.48106	-0.57125
3	1	0.2	0.75	1	1	2.67121	-1.84356	0.17235
4	1	0.2	0.75	0.8	0.81	1.92978	-2.47877	1.54898
5	0.2	0.02	0.02	0.2	0.7	2.26732	-2.347	1.07968
6	0.2	0.02	0.02	0.25	0.65	2.73861	-1.87854	0.13994
7	0.7	0.25	0.25	0.65	0.3	1.80506	-2.85776	2.05269



## Подбор оптимального сглаживания 2 момента

2 пространственный момент описывает плотность всевозможных пар особей в А

Исследование проводилось с параметрами из работы М.Л. Куркина, при этом увеличили кол-во итераций до 2000, чтобы уменьшить погрешность

Биологический параметр	Значение
$m(\xi)$	0.96
$w(\xi)$	0.09
$b$	0.2
$d$	0.02
$d'$	0.02

## Варианты: Скользящее среднее

Среднее арифметическое значений исходной функции за установленный период

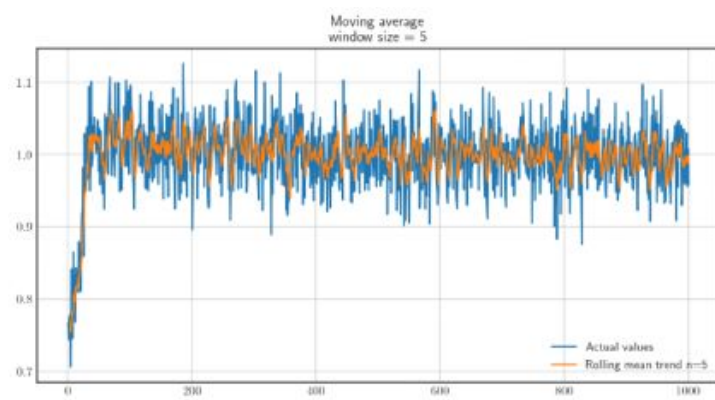
Подбор оптимального интервала: вывод - шумность графика сильно

снижается при возрастании задаваемого параметра

### Недостатки:

- Сглаживание не чувствительно к резким скачкам
- Первые  $n-1$  наблюдений не отображаются

$$\hat{y}_t = \frac{1}{n} \sum_{i=0}^{n-1} y_{t-i}$$





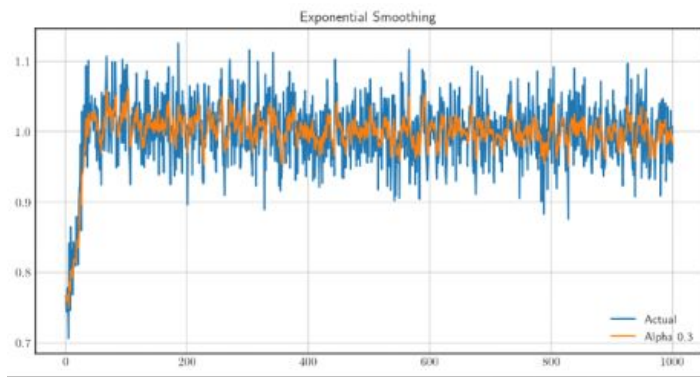
## Варианты: простое экспоненциальное сглаживание

$$\hat{y}_t = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_{t-1}$$

- При прогнозировании, более новым значениям наблюдаемой величины присваивается больший вес по сравнению с более старыми значениями, при этом более старым значениям присваиваются экспоненциально убывающие веса.
- Чем меньше  $\alpha$ , тем в большей степени фильтруется, подавляется шум, а также уменьшается влияние прошлых наблюдений.

### Недостатки:

- Невозможно оценить результат визуально





## Варианты: двойное экспоненциальное сглаживание

Используется для данных, показывающих четкую тенденцию.

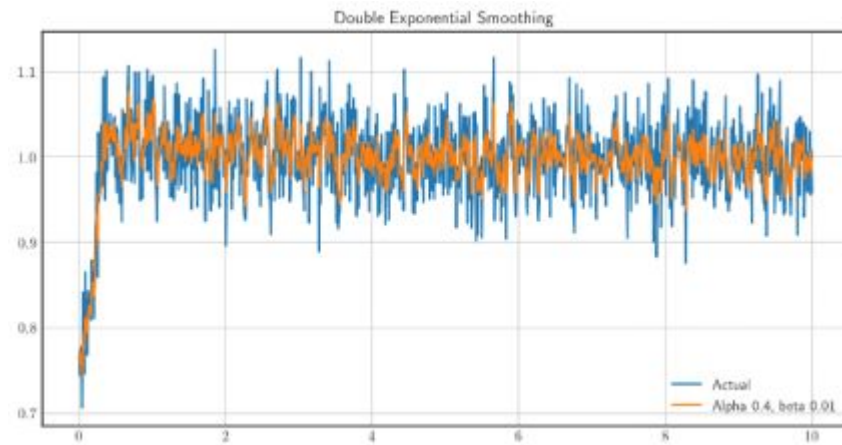
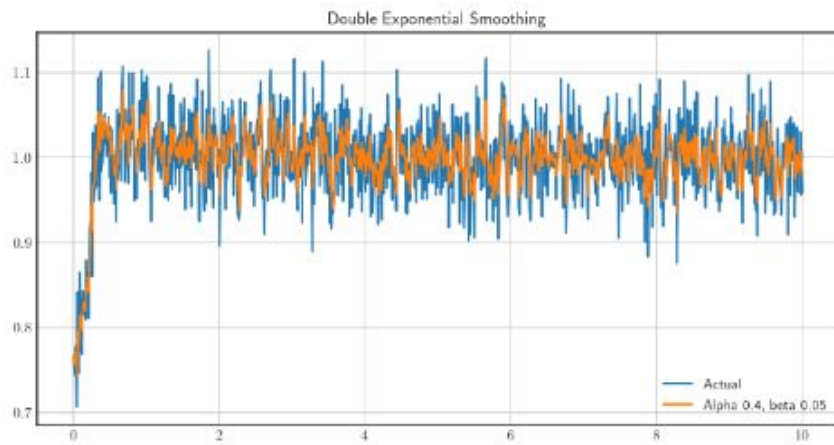
Данные разбиваются на уровень (ожидаемое значение ряда) и тренд, который изменяет вклад предыдущего значения в зависимости от тенденции. Для тренда тоже применяется экспоненциальное сглаживание, чтобы увидеть, зависят ли будущие изменения от предыдущих

$$l_x = \alpha \cdot y_x + (1 - \alpha) \cdot (l_{x-1} + b_{x-1})$$

$$b_x = \beta \cdot (l_x - l_{x-1}) + (1 - \beta) \cdot b_{x-1}$$

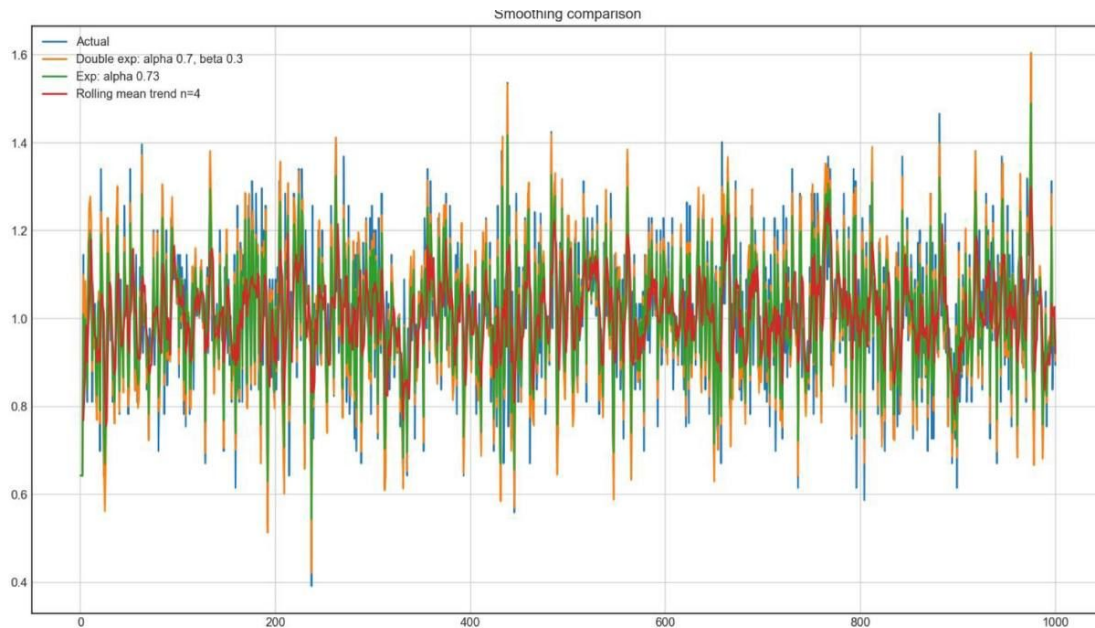
$$\hat{y}_{x+1} = l_x + b_x$$

## Варианты: двойное экспоненциальное сглаживание



При минимальном параметре  $\beta$  сглаживание проявляет себя лучше. Поэтому предыдущие изменения не влияют/влияют минимально на последующие изменения.

## Оценка сглаживаний:



Визуально лучшим  
кажется скользящее  
среднее



## Оценка сглаживний: метрики

Для оценки использовались метрики:

- Дисперсия отрезков длины 100
- Среднеквадратичная ошибка отрезков длины 100
- Среднеквадратичная ошибка всех наблюдений

В итоге выбрано простое экспоненциальное сглаживание, показавшее себя лучше, чем скользящее среднее на данных метриках. Выбранный параметр 0.45



## 4. ПОИСК ТОЧКИ ОСТАНОВА

Необходимо определить по численности популяции определить выход на плато

Для исследования особенностей поведений популяций был выбрано простое экспоненциальное сглаживание с ( $\alpha=0.45$ ). Сглаживание применялось последовательно 5 раз

## Пример сглаживания

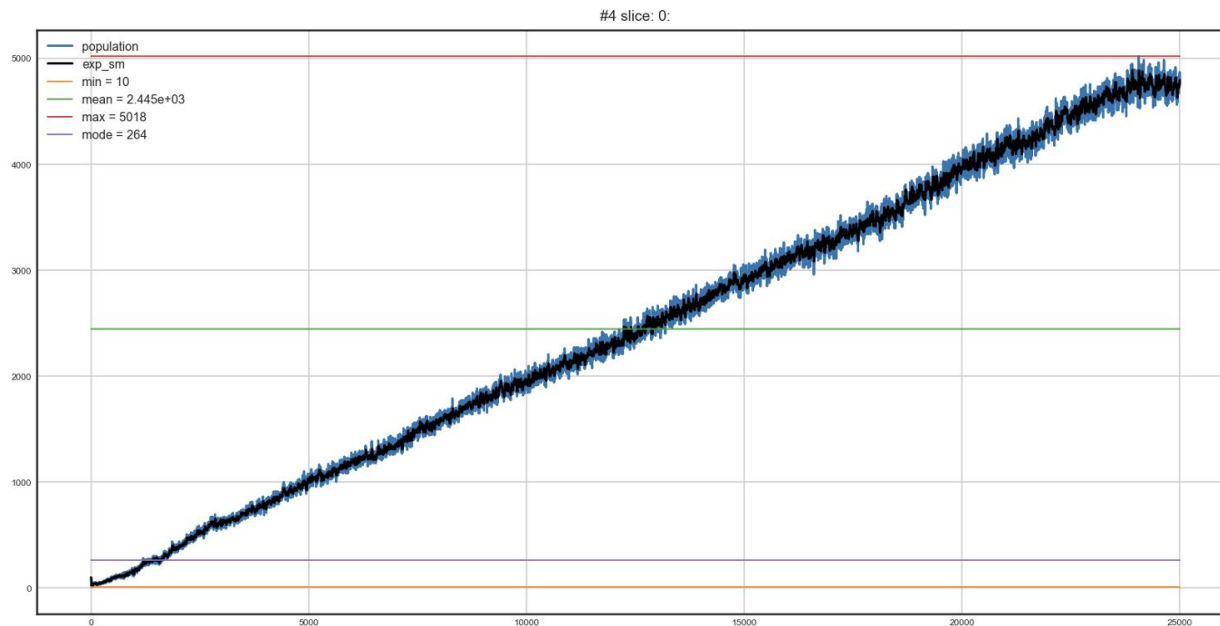
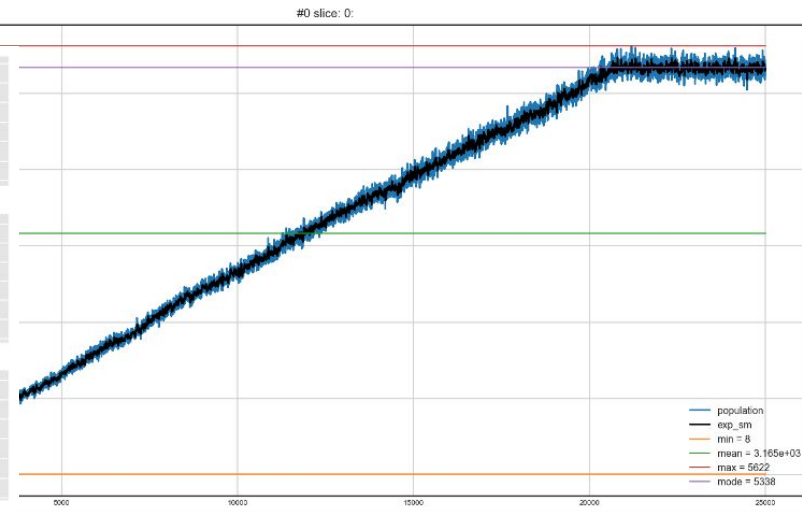
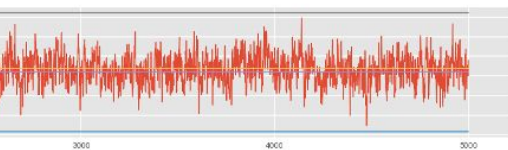
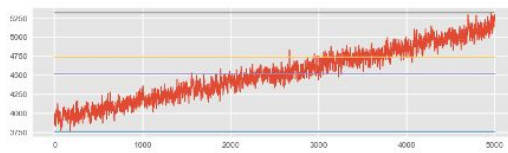
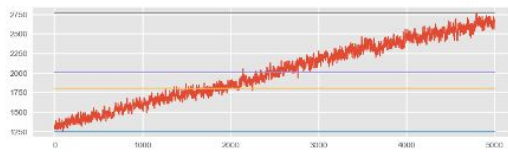
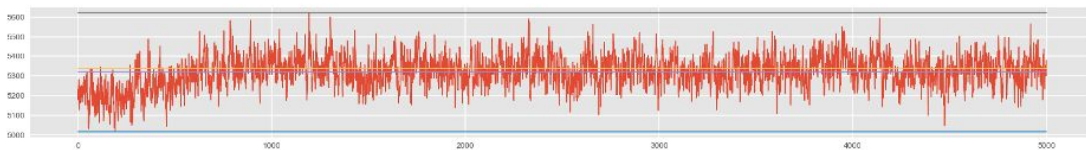
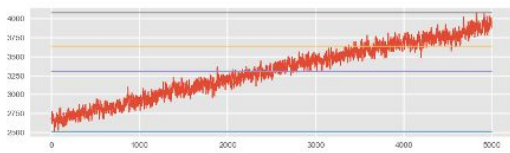
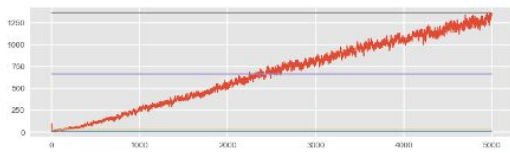


Рисунок 16 – Пятое последовательное простое эксп. сглаживание на популяции №5

# Особенности графиков

Графики разбили на 5 частей и рассмотрели минимум, максимум, среднее и моду для всего графика и каждого участка отдельно







## Выводы:

Шум растет с ростом числа индивидов в популяции

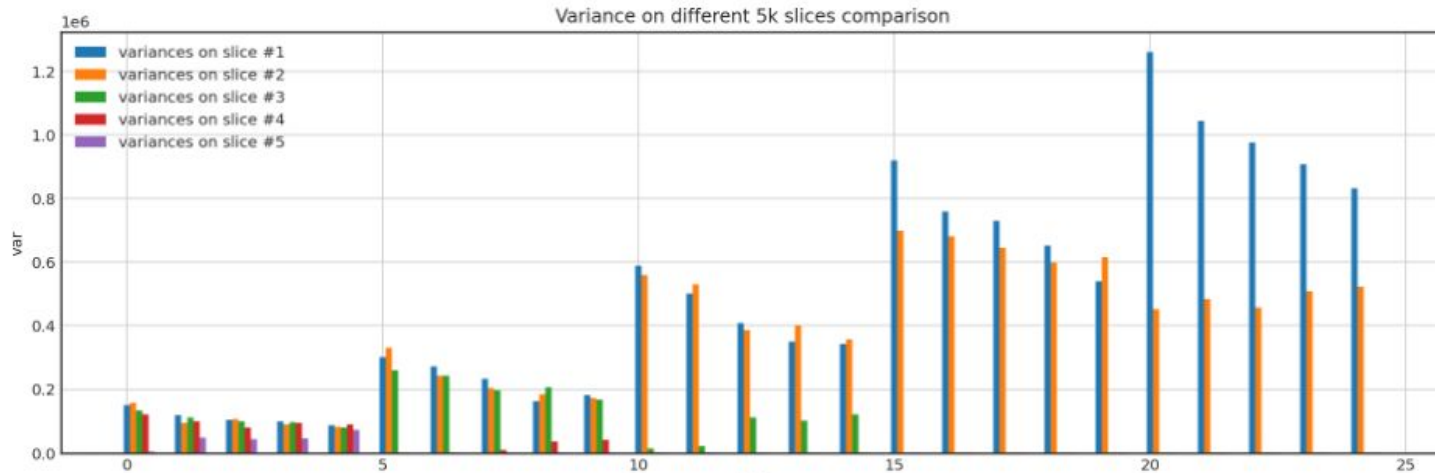
Значения среднего и моды становятся ближе с увеличением доли плато на срезе.

Дисперсия падает с выходом на плато

$$D[X] = \sum_{i=1}^n p_i \cdot (x - E[X])^2, \text{ где } E - \text{матожидание.}$$

# Выделение групп популяций по дисперсии

1. Группа графиков №№1-5 – плохо поддаются анализу, различия дисперсий разных срезов выражены слабее всего.
2. Группа графиков №№6-10 – не очень явно выражены различия дисперсий срезов.
3. Группа графиков №№10-15 – хорошо поддаются анализу, различия дисперсий разных срезов выражены достаточно заметно.
4. Группа графиков №№15-25 – отлично поддаются анализу, различия дисперсий разных срезов выражены однозначно сильно.



## Группа 2

Наблюдаем высокий показатель первого среза в сравнении с минимальным на последнем, пятом, срезе. То, что дисперсия первого среза не является абсолютным максимумом среди всех дисперсий срезов для каждой популяции, объясняется не только стохастичностью получаемых при симуляциях данных, но и характерными симулируемым в модели Дикманна и Лоу популяциям «скачками» численности индивидов.

Применив пять последовательных простых экспоненциальных сглаживаний (рисунок 32), мы можем привести пропорции дисперсий к еще большему визуальному различию. Снижение шума самого графика популяции, позволяет получить более точно характеризующую поведение популяции информацию.

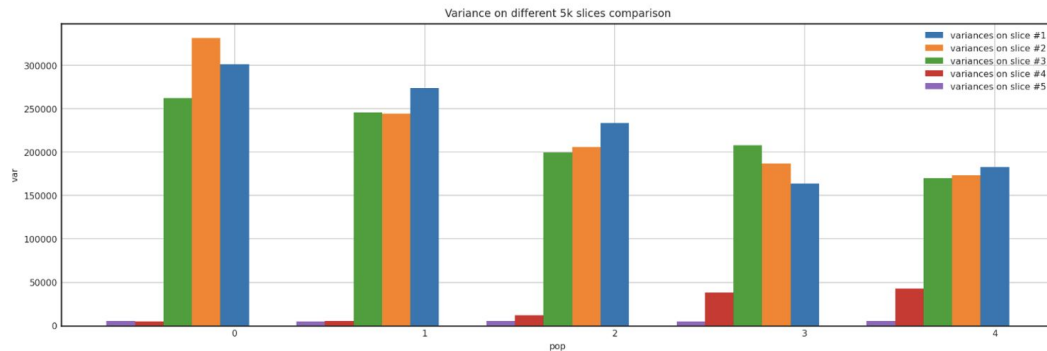


Рисунок 31 – Сравнение дисперсий для срезов по 5000, группа 2

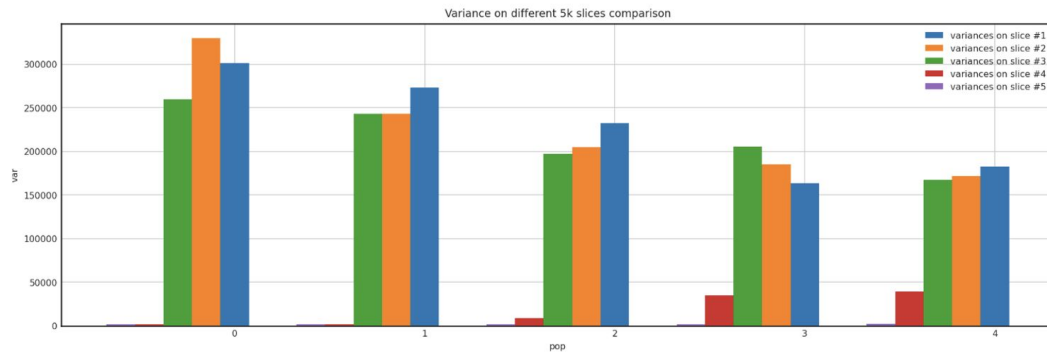


Рисунок 32 – Сравнение дисперсий для срезов по 5000, пятое последовательное сглаживание популяций группы 2

# Группа 3



Группа 3 (рис. 33–34). Можно заметить аналогичную картину с дисперсиями срезов популяций группы 2. Однако минимальные значения дисперсий отрезков для всех популяций группы теперь видны не только на последнем, но и на предшествующем ему срезах.

Применение сглаживаний (рисунок 34) также предоставляет результат более очевиден.

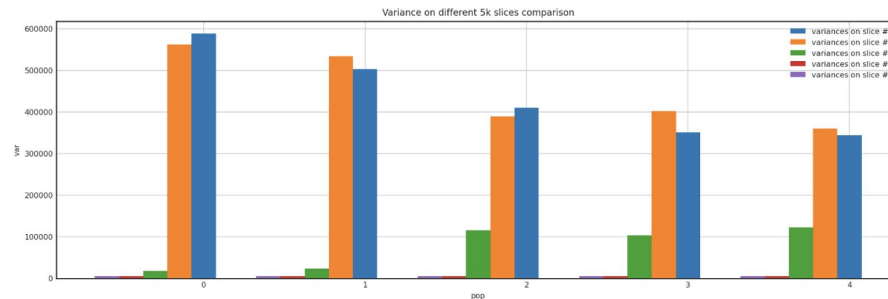


Рисунок 33 – Сравнение дисперсий для срезов по 5000, группа 3

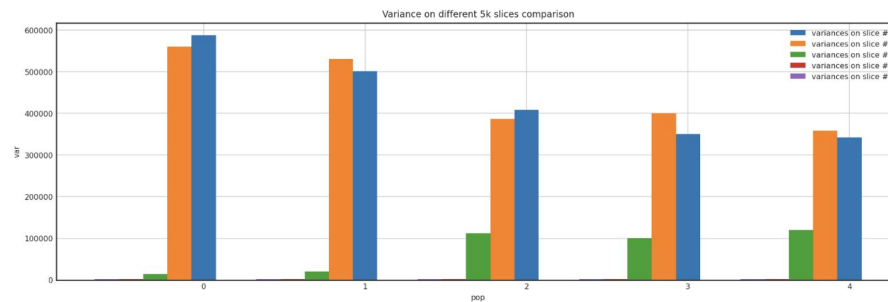


Рисунок 34 – Сравнение дисперсий для срезов по 5000, пятое последовательное сглаживание популяций группы 3

# Группа 1



Поскольку для группы 1 анализ затруднителен, подберем меньший размер среза, чтобы не наблюдать погрешности, порождаемой размером среза 5000 относительно плато популяции №5 (как указывалось в пп. 4.1, плато содержит 1200 наблюдений). Для этого сократим размер срезов до 2500 и до 1000 наблюдений на срез

Отметим, что такое сокращение срезов (2500 наблюдений, рис. 37–38), позволяет увидеть описанные выше различия первого и последнего срезов более явно, в частности для популяции №5.

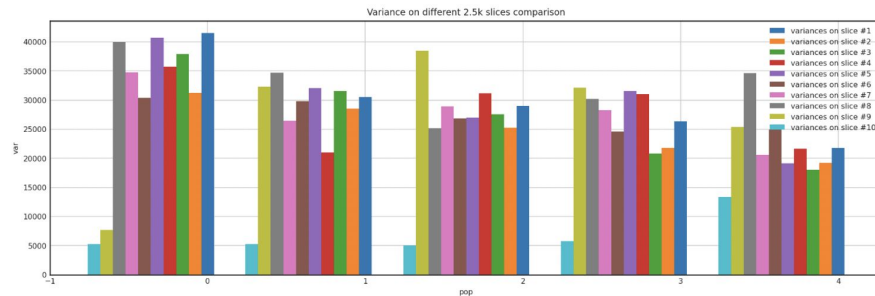


Рисунок 37 – Сравнение дисперсий для срезов по 2500, группа 1

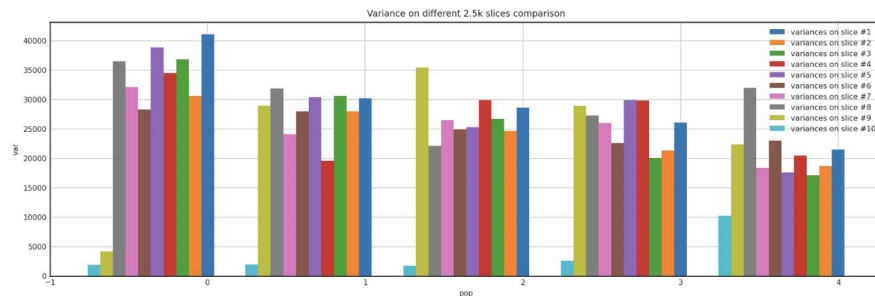


Рисунок 38 – Сравнение дисперсий для срезов по 2500, пятое последовательное сглаживание популяций группы 1

# Группа 1



Сокращение срезов до 1000 наблюдений (рис. 39–40), которое, предположительно, могло лучше охарактеризовать срезы для популяции №5, подводит нас к выводу о том, что излишне короткие срезы помешают увидеть описанные в пп. 4.3.1 свойства дисперсии на различных участках графика численности популяции, поскольку с сокращением размера среза, начинают приобретать больший вес стохастичные колебания численности популяции, и меньший вес получает общий тренд поданного участка графика численности популяции.

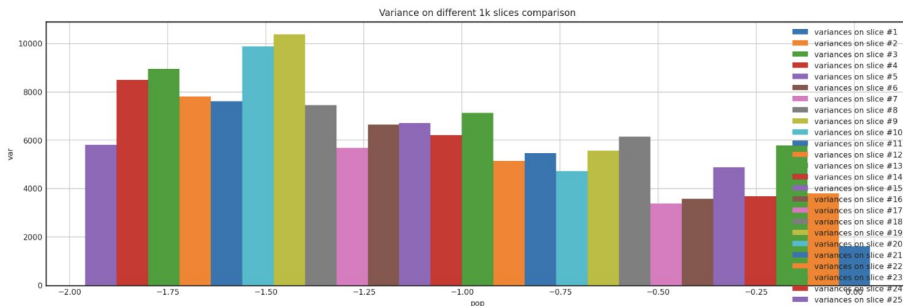


Рисунок 39 – Сравнение дисперсий для срезов по 1000, популяция №5

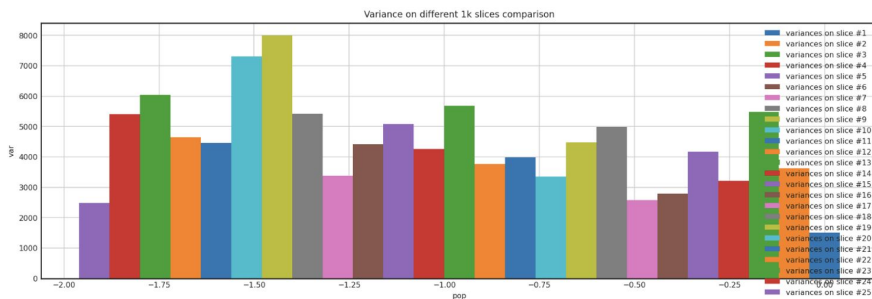


Рисунок 40 – Сравнение дисперсий для срезов по 1000, пятое последовательное сглаживание популяции №5

# Группа 4

Для популяции №15 срезы по 1000 наблюдений показывают хороший результат относительно разницы дисперсий первого и последнего срезов, однако такая точность будет излишней, поскольку уже на срезах по 5000 наблюдалась тенденция, которую мы стремимся выделить, описать и формализовать..

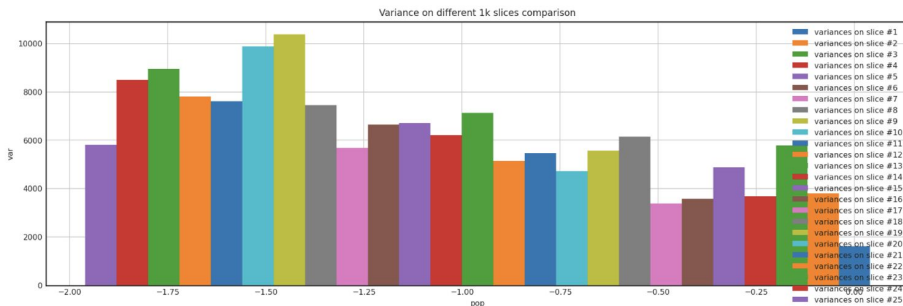


Рисунок 39 – Сравнение дисперсий для срезов по 1000, популяция №5

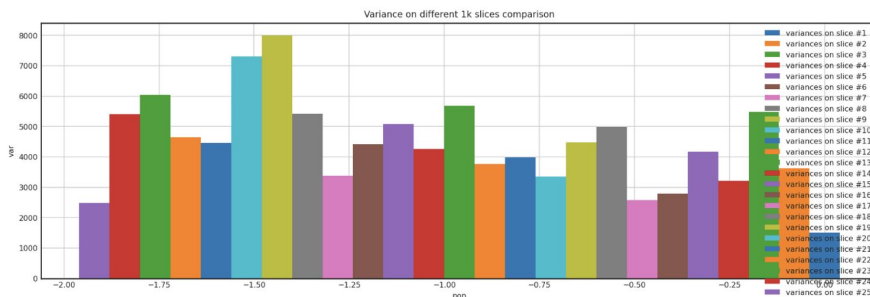


Рисунок 40 – Сравнение дисперсий для срезов по 1000, пятое последовательное сглаживание популяции №5



## Вывод

Итак, подводя итог детальному рассмотрению дисперсий и выделению групп популяций, можно сказать, что участки графиков численности популяций достаточно точно описываются срезами по 5000 и по 2500. В процентном соотношении размеры срезов составляют 20 и 10 процентов от общего количества наблюдений (25000, пп. 4.1) соответственно.





## Группа 4

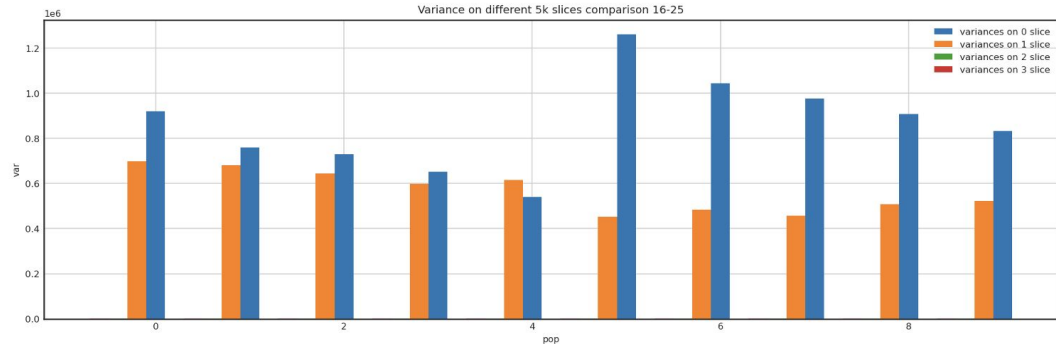


Рисунок 43 – Сравнение дисперсий для 1–4 срезов по 5000, пятое последовательное сглаживание популяций №16–25

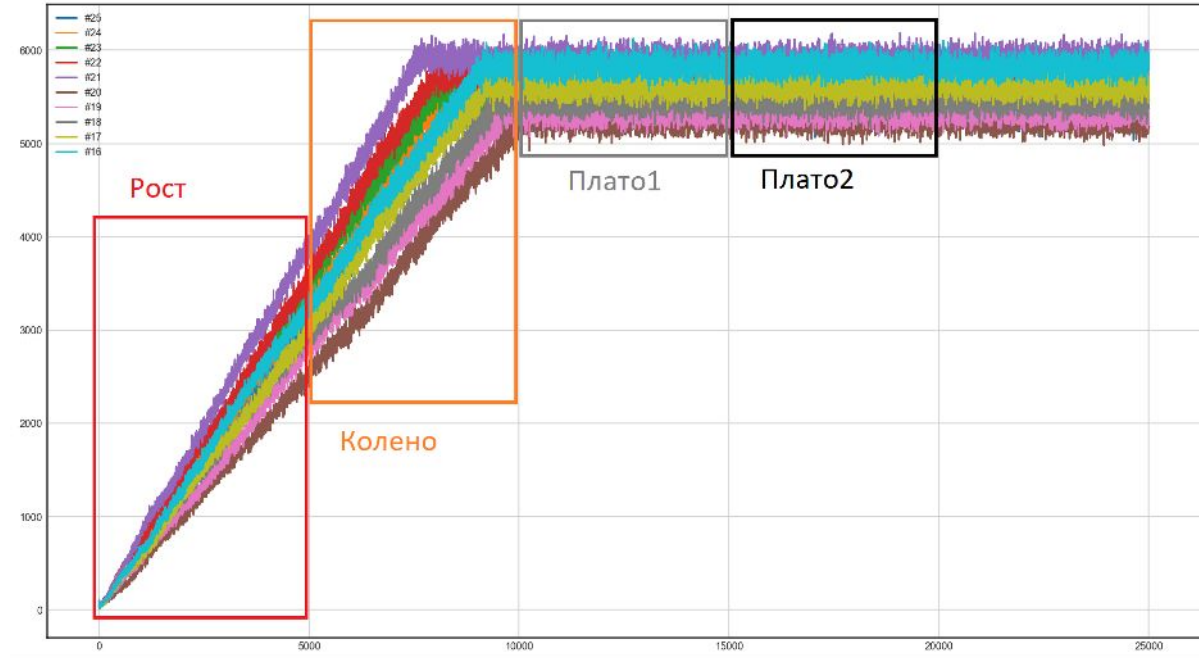
Выберем срезы 1–4 для популяций группы 4 (пп. 4.3.2). Рассмотрев рисунок 43, опишем наблюдения, максимальная дисперсия наблюдается на первом срезе, для третьего и четвертого срезов она примерно равна и составляет не более 1750 для популяций №№16–25 (против для первых срезов).

В популяциях №№16–20 на втором срезе можно видеть значения дисперсии, сопоставимые таковыми для первого среза этих популяций. В популяциях №№21–25 дисперсия вторых срезов принимает промежуточное положение между минимальными 3–4 и максимальными первыми. Позже эти различия будут подробно объяснены.

# Группа 4

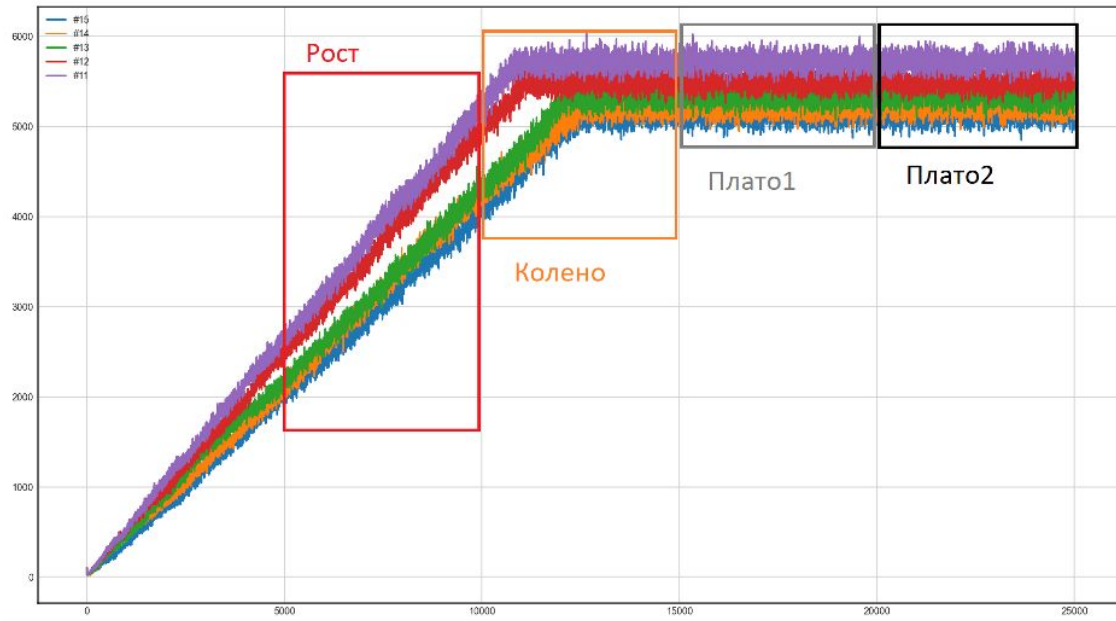
Если наложить наблюдаемые дисперсии (рис. 43) на графики численностей популяций №№16-25 и дать им условные названия (рис. 44), мы получим визуальную интерпретацию данных, анализированных в пп. 4.3.1–4.3.2):

1. Участок «Рост» – характеризуется высокой дисперсией, значение которой будет зависеть от скорости выхода популяции на плато. Соответствует первому срезу на рис. 43.
2. Участки «Плато1», «Плато2» – участки, наиболее близкие по значению численности популяции к константе (однако не равные ей вследствие стохастичности численностей популяций в биологической модели Дикманна и Лоу). Соответствуют срезам 3–4 на рис. 43.
3. Участок «Колено» – участки спорной дисперсии, которую невозможно однозначно отнести к «Росту» или «Плато». Соответствуют второму срезу на рис. 43.



## Группа 3

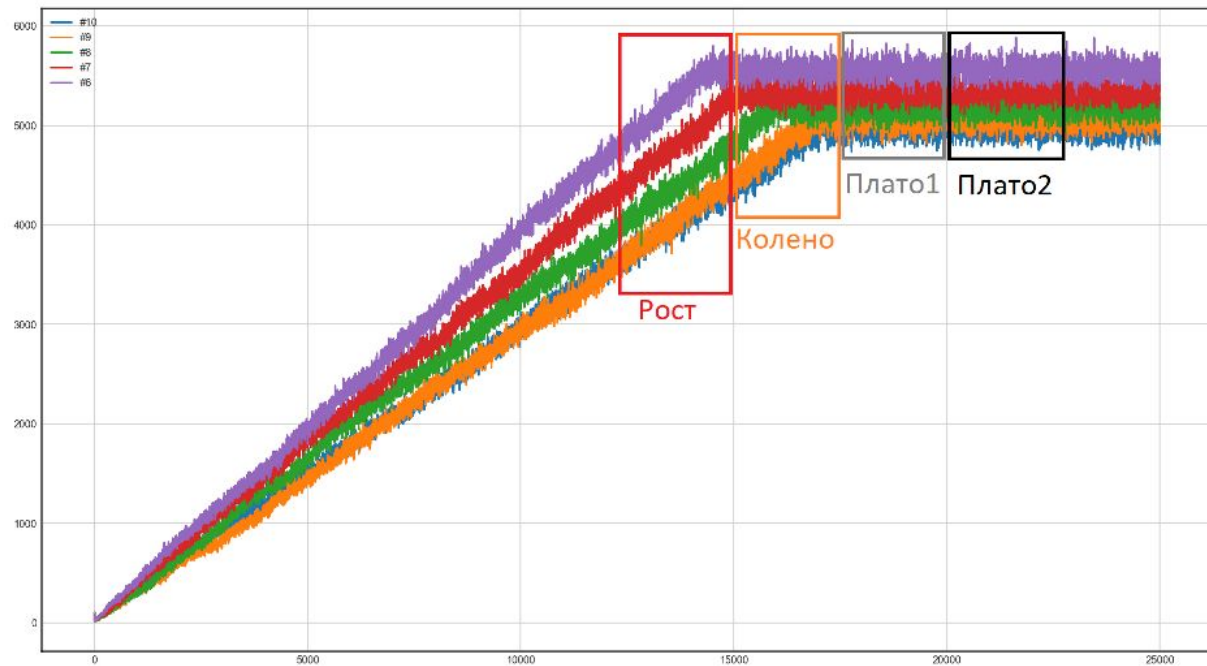
По аналогии с рассмотренной группой 4, проанализируем группу 3 (рис. 45). Выберем те же срезы, в силу схожести их пропорций с аналогичными в группе 4. Заметно меньшее (относительно первого среза, «Роста») значение дисперсии на срезе «Колена» численностей популяций группы 3.



## Группа 2

Рассмотрим группу 2 (рис. 47), для большей точности, выберем срезы по 2500 наблюдений, по причинам, описанным в пп. 4.3.2.

Дисперсии срезов «Плато1» и «Плато2» относительно дисперсии среза «Роста» заметно поднялись (если сравнивать с группами 3, рис. 45, и 4, рис. 43), очевидно, на это влияет более пологий выход на плато и более короткие срезы в сравнении с группами 3 и 4 (см. пп. 4.1 и 4.3.2).



# Формализация характеристик

Как обсуждалось ранее (пп. 4.3.1), зона «Колена» неоднозначна, из таблицы 3 следует, что соотношения «Колена» с «Ростом» или «Плато» не стоит принимать как значение, способные охарактеризовать выход на плато.

Дополним приведенные в таблице 3 наблюдения тем, что соотношение дисперсий двух произвольных последовательных отрезков равной длины на зоне «Роста» примерно соотносятся 1 к 1.

Итак, мы можем определить зону «Плато» как два отрезка на общем графике численности популяции равной длины, соотношение которых равно примерно 100% (по таблице 3). Соотношения «Рост»-«Плато» разнятся между группами, но однозначно являются максимальными среди наблюдаемых, подкрепим определение зоны «Плато»: **отношение дисперсии предшествующего аналогичного отрезка к дисперсии первого (содержащего более ранние наблюдения, относительно второго) из названных отрезков должно быть максимальным среди подобных соотношений для всех пар таких отрезков.**

Данная формулировка не дает однозначного алгоритма выявления плато, но позволяет нам создать алгоритм, способный достаточно точно определить плато.

Таблица 3 – Соотношения участков графиков численностей популяций групп 2–4

Группа	Соотношение		
	Рост - Колено	Колено - Плато1	
2	1389%	1011%	
3	1530%	3992%	
4	160%	32801%	
Группа	Соотношение		
	Рост – Плато1	Рост – плато2	Плато1 – Плато2
2	2798%	2815%	104%
3	25189%	25202%	100%
4	50867%	52508%	104%

Группа	Соотношение		
	Рост – Колено		Колено – Плато1
2	1389%		1011%
	Рост – Плато1	Рост – Плато2	Плато1 – Плато2
	2798%	2815%	104%
	1530%		3992%
3	1530%		3992%
	Рост – Плато1	Рост – Плато2	Плато1 – Плато2
	25189%	25202%	100%
	160%		32801%
4	160%		32801%
	Рост – Плато1	Рост – Плато2	Плато1 – Плато2
	50867%	52508%	104%
	50867%		52508%



# Алгоритмы

Приложение В. На основании приведенной в пп. 4.4.2 формулировки был реализован алгоритм, соотносящий дисперсии  $i$ -того отрезка с отрезком  $i + 1$  и, в случае удовлетворения условию о соотношении участков плато, вносящий соотношение отрезка  $i - 1$  к  $i$ -тому отрезку в словарь. Результат работы алгоритма – индекс среза, для которого получено максимальное соотношение отрезка  $i - 1$  к  $i$ -тому отрезку.

Приложение С. Пересмотрим подход: отбросим параметр определения отрезков «Плато» и, чтобы избежать определения «скачков» как «Плато» на фоне возможного попадания на «Колено» в отрезке  $i - 1$  и незначительного отличия дисперсий «Колено» и «Плато» (см. пп. 4.4.1, 4.4.2), интерпретируем отношение отрезка «предшествующего первому отрезку плато» как отношение отрезка  $i - 2$  к  $i$ -тому (предполагаемому выходу на «Плато»). Алгоритм также вычислит индекс среза, предположительно, вышедшего на плато.



## Алгоритмы

Приложение С. Отбросим выбросы на каждом срезе, определим наличие роста от отрезка  $i - 1$  до  $i$ -того отрезка как соотношение арифметических средних на этих отрезках больше 80%. Скорректируем «выбор» между отношением к  $i$ -тому отрезку отрезка  $i - 1$  или  $i - 2$  через определение их количества: для срезов маленького размера будем смотреть на  $i - 2$ , иначе – на  $i - 1$ .

Помимо индекса среза, вышедшего на плато алгоритм будет возвращать и значение отношения отрезка «Роста» к «Плато1». Также, благодаря введению порога доверия, добавляется возможность неопределения плато, в таком случае алгоритм вернет  $-1$  для обоих значений.