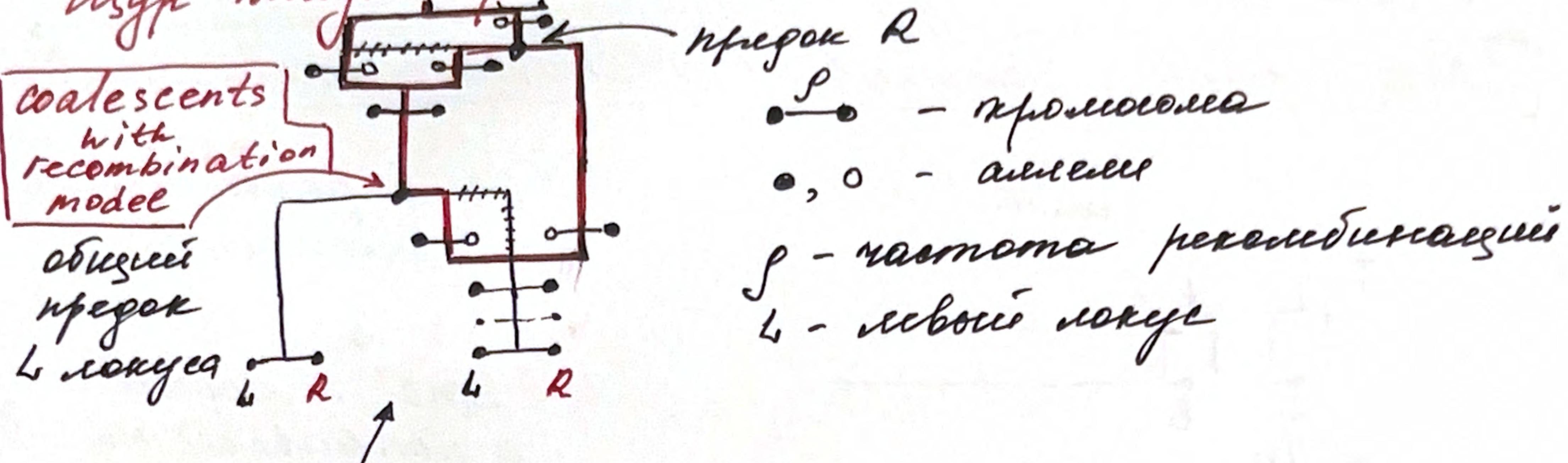


Лекция 5. Структура Марковского процесса

11 час
20.25.2023

• Чур Владислав Любовьев

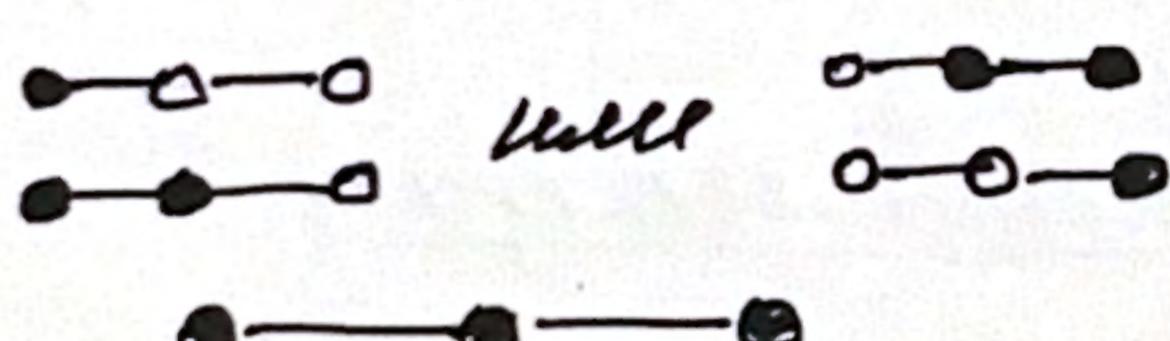


Структурный график рекомбинации
(Ancestral recombination graph)

Directed acyclic graph

точка рекомбинации

Примеч.:



содержит историю генетической линии с уверенностью, как одинаковы идентичные генотипы в популяции

! Если рассматривается 1 ложе, то не учитывается перекрёст, а не циклический график

Частота коалесценции $\frac{1}{N_e}$ числа предков

число полуподвигов $\frac{1}{N_e} K$

$$\frac{\binom{K}{2}}{N_e} \sim K^2$$



частота скрещивания

$\frac{1}{N_e}$ — вероятность коалесценции у одной пары образцов

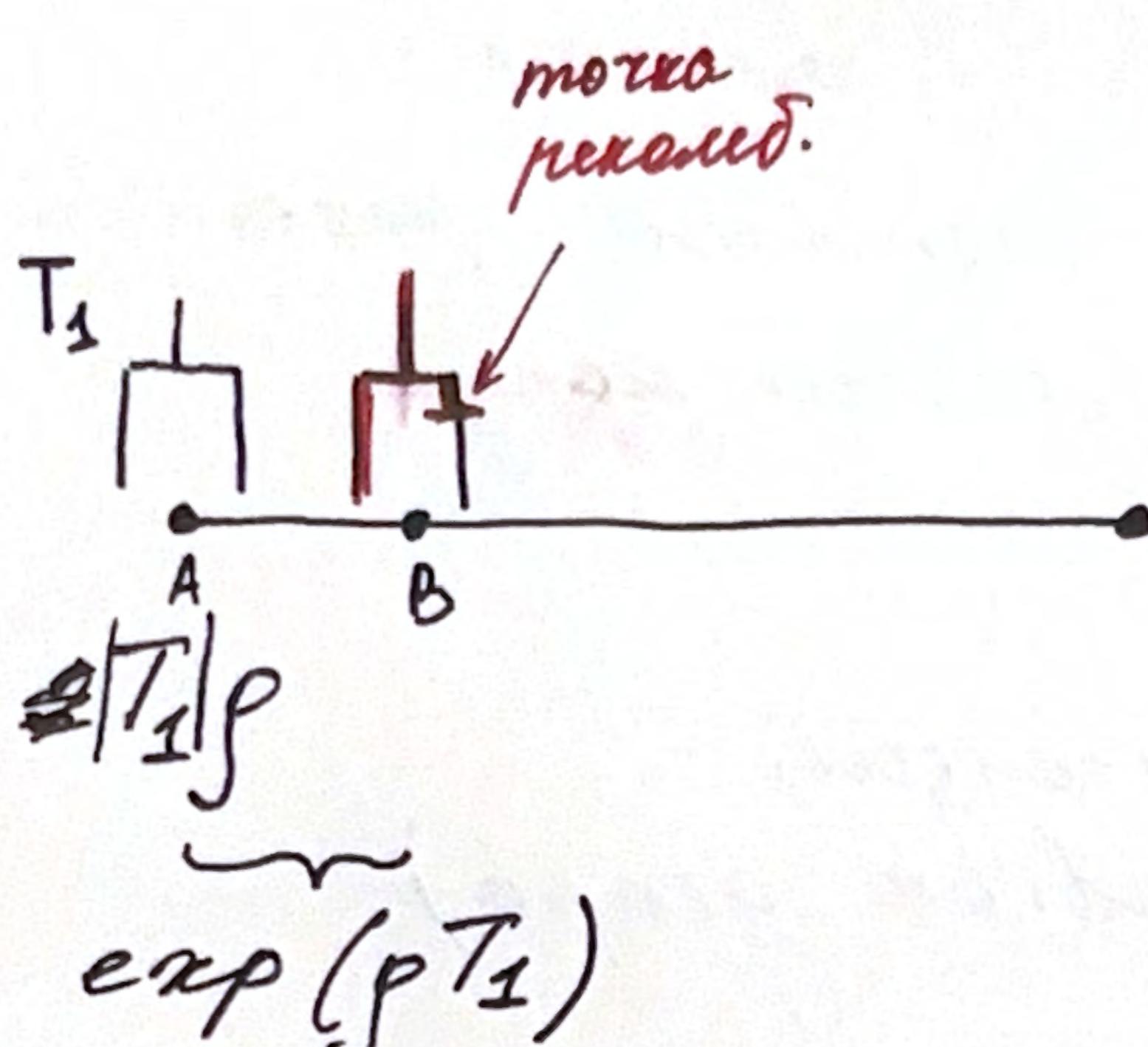
* N_e — корректированный размер популяции \rightarrow размер популяции, вид овой популяции, которая определяет потенциальную

! Для марковских процессов частота коалесценции не имеет складывания

- gK — К раз по времени процесса (частота рекомбинации).

Представление графа, как все gefieb в на хромосому:
 • Как проходим с gefieb на gefieb?

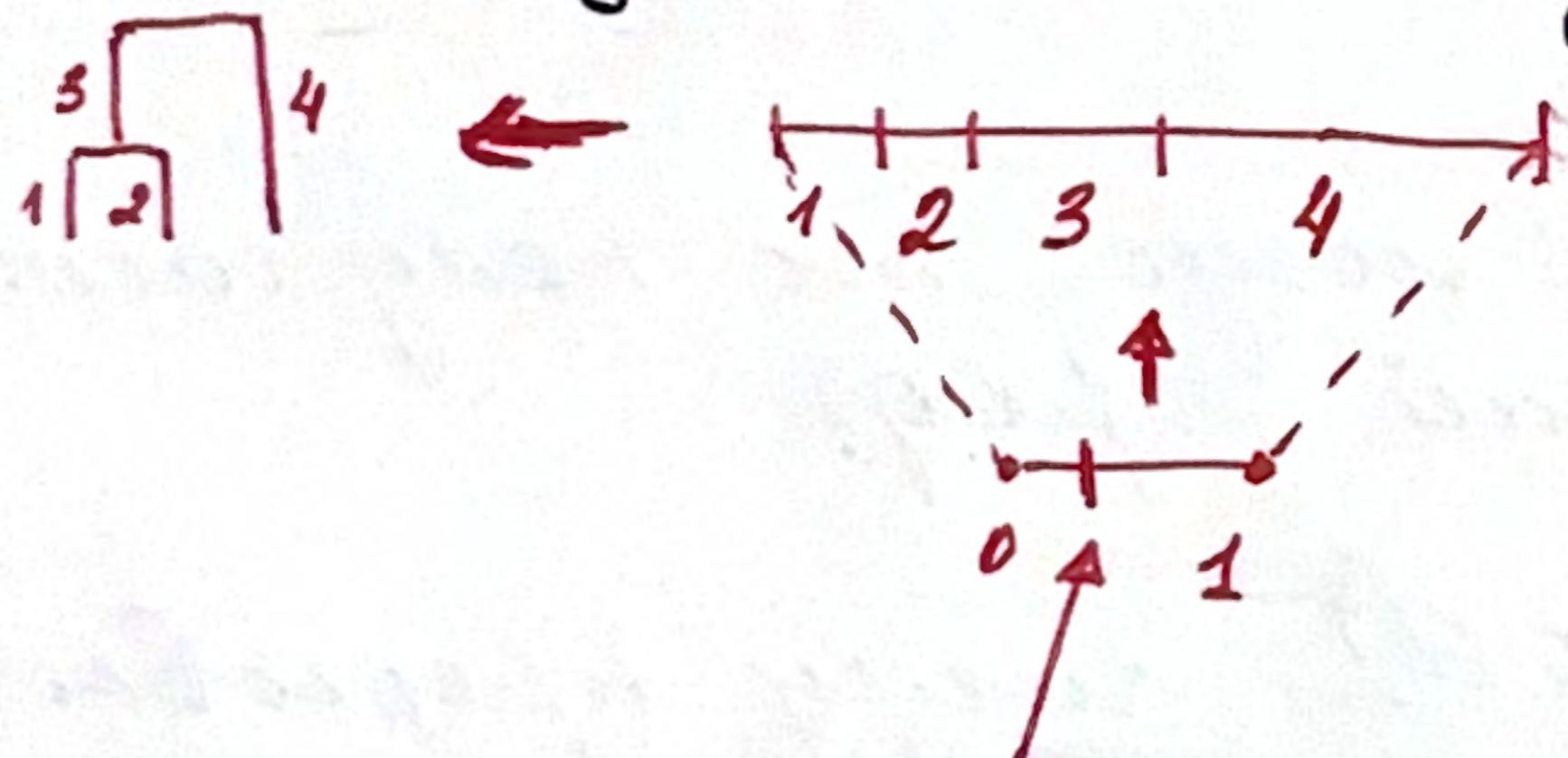
[SMC model - Sequential Moranian Coalescent]



↑ LD - коррелированное изменение
 ↓
 SMC' - модель с nonLD на LD.

$|T_1|$ - суммарная длина всех gefieb
 $|T_1|p$ - частота (исчезновение) Марковского процесса
 Женератор. процес. с этой частотой
 идет от море B к морю A

{ Как определить моря разделяющие ?

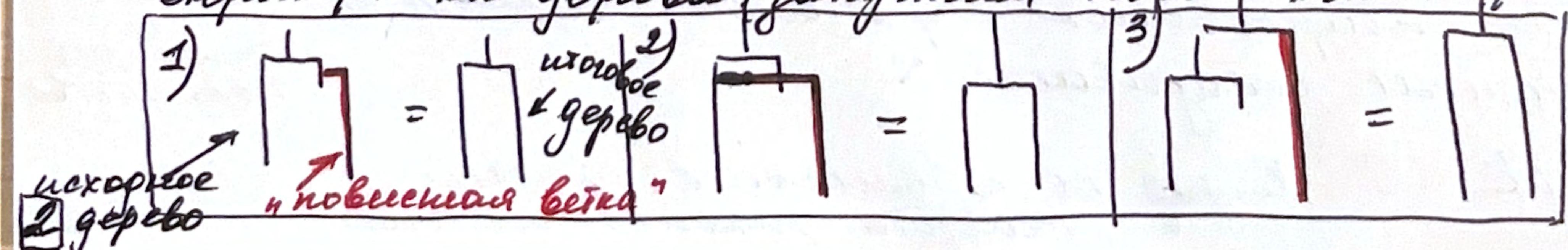


предсказ. разделяет моря на
 отдельные и единственные на
 gefieb моря разделяющие

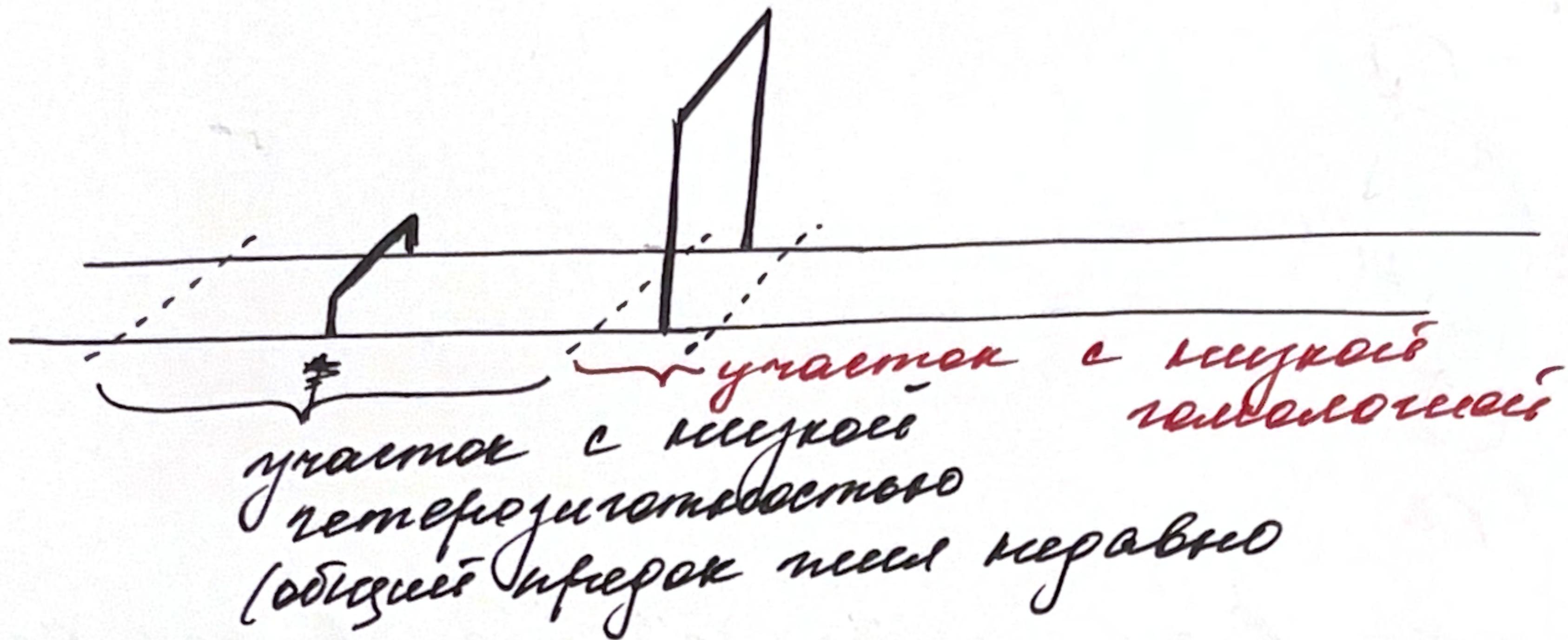
• Разделение разделяющих:

$$\begin{cases} \text{частота коалесценции} = \frac{2}{N_e} \\ \text{частота коалесценции} = \frac{1}{N_e} \end{cases}$$

Технология gefieba (запутанный новый коал. процесс):



Типы звуков с динамическими звуками: (—)



1) Распределение времени во времени

$N_e(t) \rightarrow$ из 1) видно, как меняется количество новых поглощений

$\frac{K(t)}{N_e(t)}$ → нелинейный закон изменения

Hidden Markov Model

• N_e - время конфигурации звуков в некоторое

предсказуемое будущее

Пример:

• атмосферное давление (не видим):

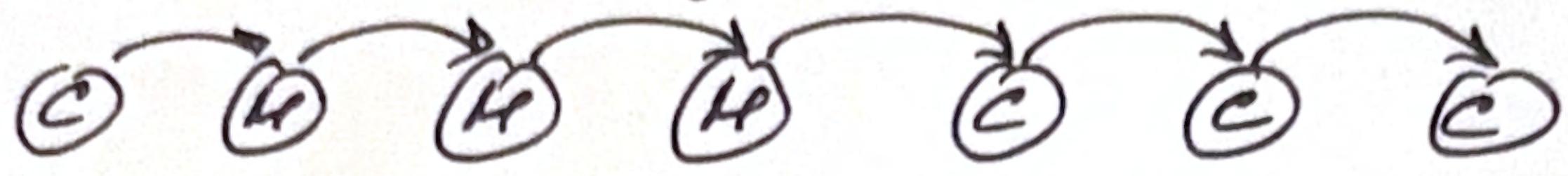
Ⓐ - cold

Ⓑ - hot

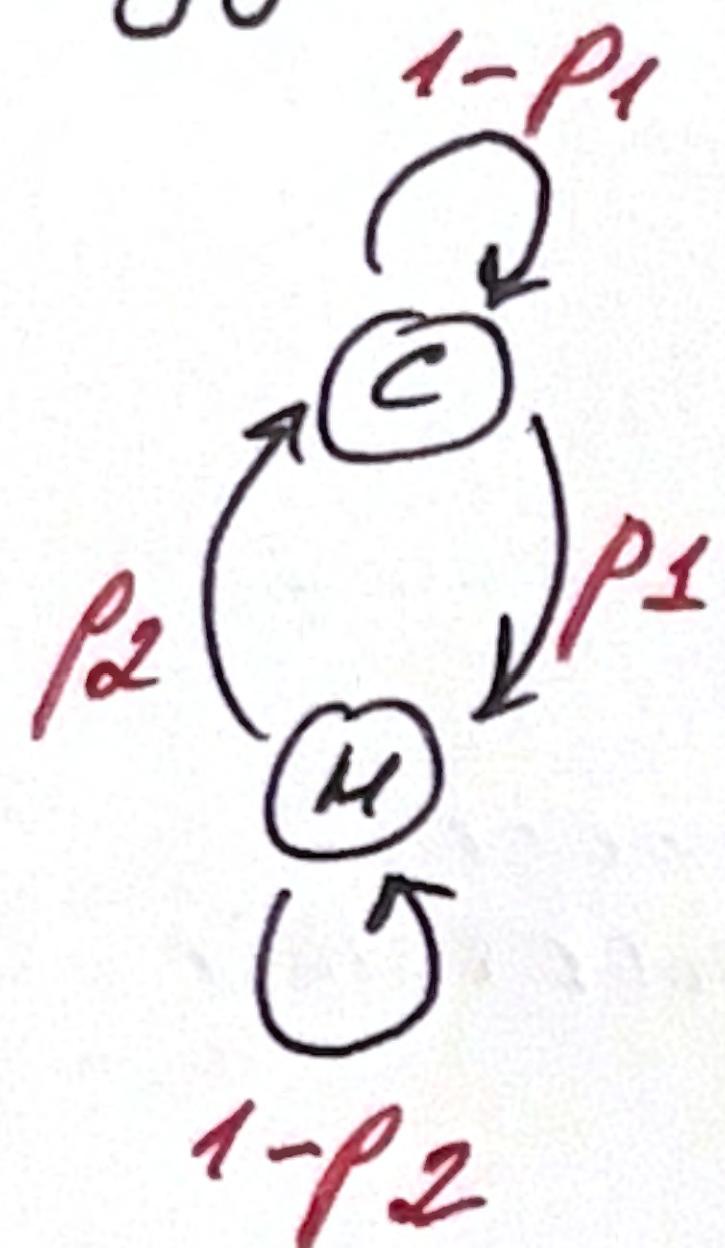
• наблюдения:



дневацца бабулике:



Какова вероятность, что за годовщину гибели последует менингит? и т.д. . .

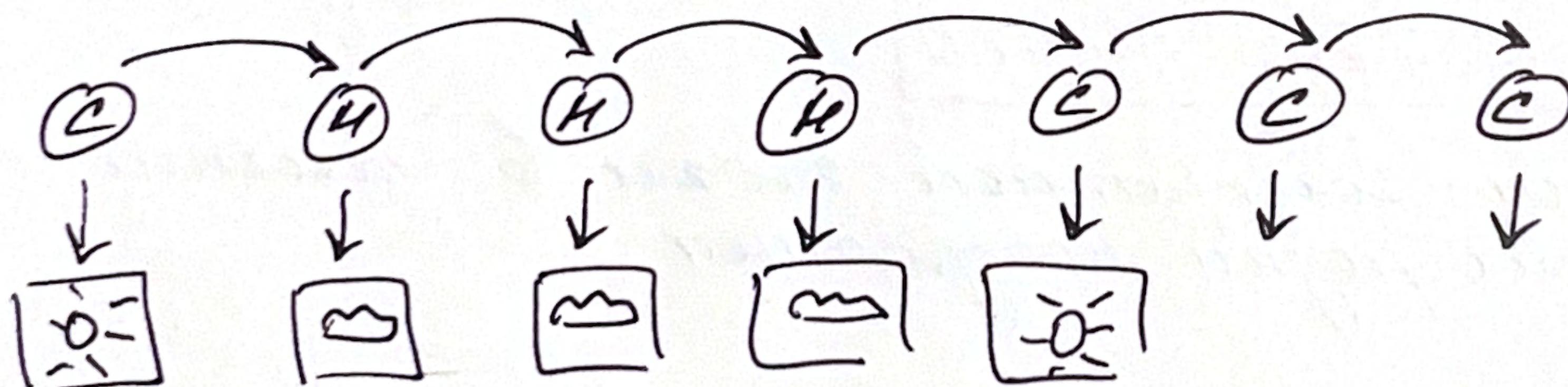


Какова вероятность смерти в год. гибели?

$$P(\text{死}/\text{C}) = \frac{c, \text{死}}{c}$$

→ Матрица перехода:

$$A = \begin{pmatrix} 1-p_2 & p_1 \\ p_2 & 1-p_2 \end{pmatrix}$$



! Сифилитическое сопутствующее холода воспаление
(заболевание по последовательности наблюдений)

! В геноме

1) Чадогорение - напарное симптомное
братья погибают

2) Сифилитическое восп. - выражена по общему
признаку

Число параметров для языка HMM (матр. скрытого):

1) A - transition matrix

$$A = (P(s_i | s_j))$$

N = # of hidden states

i, j - скрытые состояния

A - $\underbrace{N \times N}_{\text{матрица}}$

possible

2) E - emission matrix
 K = # возможных наблюдений (эмиссий)

2) E - emission matrix

$$E = (P(e | s))$$

$\uparrow \quad \nwarrow$
источник
наблюдение

E - $\underbrace{N \times K}_{\text{матрица}}$

3) B - beginning distribution (начальное распределение)

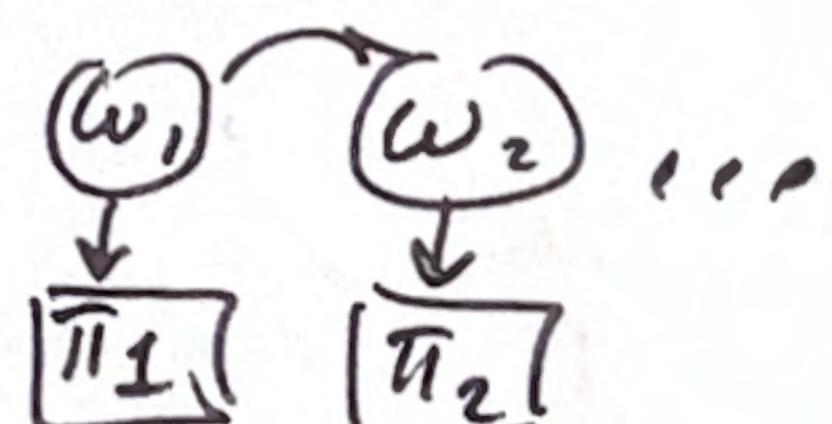
$$B = (P_0(s))$$

м.е. с какой P в стартовом состоянии начнется
последовательность слов? ($\Pi_{\text{старт}}$)

...

(I) Доминированием наблюдений

w_1, \dots, w_L
Что хуже?



a) Для A, E, B найти наилучшее описание
каких либо информативных состояний

$\hat{\pi} = \underset{\pi}{\operatorname{argmax}} P(\pi, w)$ - Viterbi algorithm
наиболее

вероятной $\hat{\pi}$ - каких либо информативных состояний
 $\hat{\pi} = (\hat{\pi}_1, \dots, \hat{\pi}_L)$ $\hat{\pi}_i$ - i -е самое вероятное из 5

5) Часто сопоставляют $P(w)$

→ априорицеское предсказание об обратном ходе

6) Часто говорят о постепенности вероятности, что на каждом шаге это более и более вероятно (5) при условии бесконечной последовательности (6)?

$$P(\pi_i = s | w)$$

→ демонстрирует норму непрерывности отклонений

из 5)

7) Стартовое нач. условие - это выбор M_L .

→ Априорицеское байеса - Учебка

(правило начального условия EM - априористич., как в k-means)

↓

из каждого обновления все параметры меняются

* Значение байеса и начальному считаются, суммаем E

* А также байеса

* Тогда E параллелизируется B.

I. Априористический Витебский

? Каждое след. состоян. зависит от предыдущего
байес. условие

$$P(\pi_i, w) = P_0(\pi_{i1}) \cdot P(w_1 | \pi_{i1}) \cdot \underbrace{P(\pi_{i2} | \pi_{i1})}_{\text{из } \pi_1 \text{ попад в } \pi_2} \cdot P(w_2 | \pi_{i2}) \cdots \\ \cdots \cdot P(\pi_{iL} | \pi_{i,L-1}) \cdot P(w_L | \pi_{iL})$$

N - багнинское состояние ($\text{погодное}/\text{мёдное} = 2$)
L - число переходов между состояниями и б. последоват.
(самое частое среди на
= 30 \rightarrow картина)

N^L - все возможные состояния за все времена

2^{30} - общ. число \Rightarrow очень большое количество возможных переходов

1) Генерация вероятностей $N \times L$:

$$N \left\{ \begin{array}{l} P_0(s_1) \cdot P(w_1/s_1) \xrightarrow{P(s_1/s_1)} \boxed{\text{проверка}} = v_2(s_1) \\ P_0(s_2) \cdot P(w_1/s_2) \xrightarrow{P(s_2/s_2)} \\ \vdots \\ P_0(s_N) \cdot P(w_1/s_N) \xrightarrow{P(s_N/s_N)} \end{array} \right. \quad \begin{array}{l} \downarrow \text{Mar 1} \downarrow \quad \downarrow \text{Mar 2} \downarrow \quad \downarrow \text{Mar 3} \downarrow \quad \downarrow \text{Mar 4} \downarrow \\ \boxed{v_2(s_1)} \quad \boxed{v_2(s_2)} \quad \boxed{v_2(s_3)} \quad \boxed{v_2(s_4)} \end{array}$$

Пояснение:

$P(s_1/s_N)$ - переход из состояния s_N в s_1

2) Задача о том что вероятность:

$$\left| \begin{array}{l} P_0(s_1) \cdot P(w_1/s_1) \cdot P(s_1/s_1) \\ P_0(s_2) \cdot P(w_1/s_2) \cdot P(s_2/s_2) \\ \vdots \end{array} \right| \max = P(w_2/s_1) = v_2(s_1)$$

от "выборки")

\hookrightarrow ищем \max , т.е. наибольшую вероятность,
и этот этот \max попадает в вероятность

• Какие из переходов дают этот \max ?

3) Чему равен временной шаг перехода при просмотре б
т₂(s₁) строки? Сформулируйте

	1	2	4
N	:	ptr = argmax V ₂ (s ₁)	
		как оптимальным предыдущим считать	

! Доказать 2) и 3) для трех способов б 1)

4) Проверить не таблица 3) и напорядок
наиболее возможной
нуль

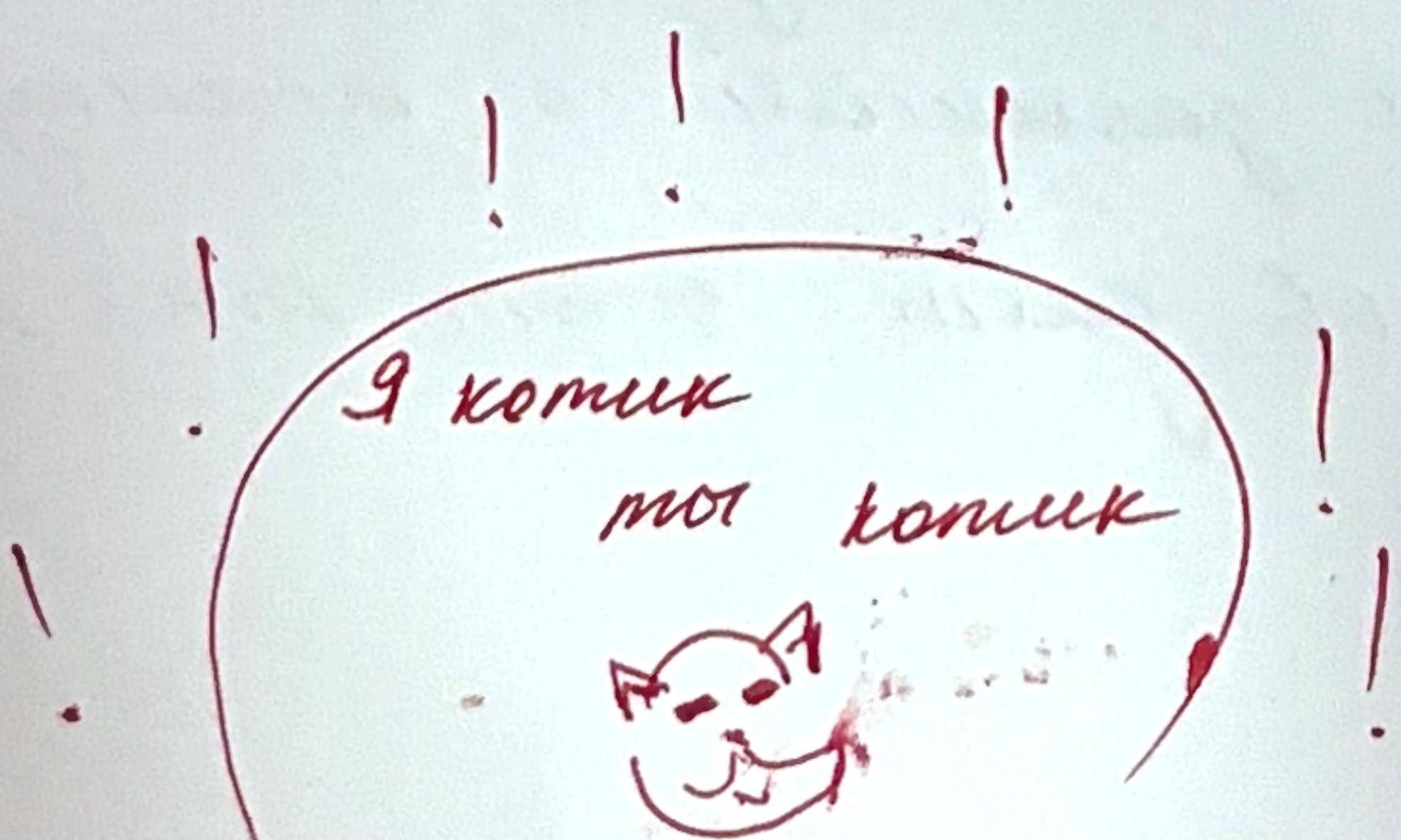


нуль по коэффициентам

! $\Theta(N^2 \cdot L_1)$ - ~~свободные от присущих к н. к. есть еще~~
~~коэффициентов~~ - сложность алгоритма ~~без учета~~ $O(N^2 \cdot L_1)$
направлена в), 3)
↑ повторяет L_1 засечек (3))

- б т₂(s₁) есть N нулей где перехода
из первого способа
- во втором способе N засечек
- всего L_1 засечек в способах

! HD! типич. ор. балансах
 $L_1 \sim O(N^2 \cdot L_1)$.



II. Доказательство вероятности последовательности наблюдений

$P(w)$

$$P(w) = \sum_{\pi} P(\pi, w)$$

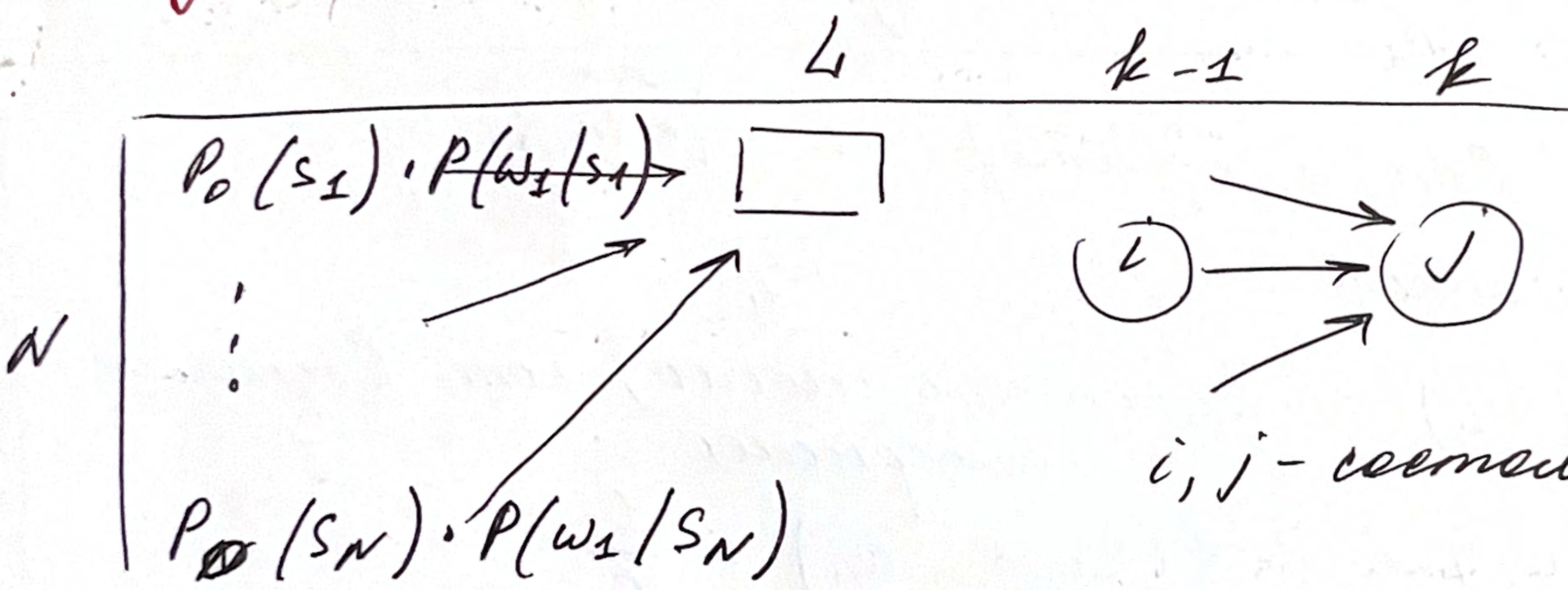
[наблюдение
пункт]

! Три неправил. Вероятность π входит в
sigma момент будем $\theta \Rightarrow$ неизвестные скрытые

$$\ln P(\pi, w) = \sum \ln P(\pi_j | \pi_{j-1}) + \sum P(w_j | \pi_j)$$

$\boxed{\ln(a \cdot b) = \ln(a) + \ln(b)}$

1) Логарифм шансов



2) Запись архивы $\boxed{\quad}$

$$P(w_k | s_i) \sum_{i=1}^N P_0(s_i) \cdot P(w_1 | s_i) \cdot P(s_1 | s_i) = P(\bar{\pi}_2 = s, w_1, w_2) = f_{W_2}(s_1)$$

↑ вероятность на бывшем шаге и для
б'я непоследовательн., наблюдают
последовательность $w_1 \rightarrow w_2$

состояние на шаге k forward algorithm

$$f_{W_k}(s_i) = P(\pi_k = \underbrace{s_i}_{\text{наблюдение}}, \underbrace{w_1 \dots w_k}_{\text{наблюдения}}) =$$

$$= P(w_k | s_i) \sum P(\bar{\pi}_{k-1} = s_i, w_1 \dots w_{k-1}) \cdot P(s_i | s_i) =$$

$$= P(w_k, s_j) \sum f_{W_{k+1}}(s_i) \cdot P(s_i | s_j)$$

3) Появляется следующий результат (здесь матрица из 1):

$$P(\pi_L = s_1; w)$$

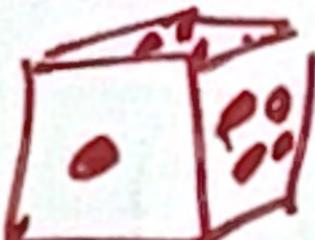
:

$$P(\pi_L = s_N; w)$$

$$\text{сумма} = P(w)$$

! O(N^2 \cdot L) - сложность алгоритма

Берется набор предложений w , и доказывается что оно входит в множество s_1

Ура! Вам и пример с :

• бросков: w_1, \dots, w_L

• $P(w_L) \leftarrow$ какое-то число, наз. вероятностью появления

$$P(w_L) = \sum_{w_1, \dots, w_{L-1}} P(w_1, \dots, w_L) = \frac{1}{6}$$

Smile! the worst is yet to come ...!

Ура! Снова! * Каждое появление нового предложения

$$P(\pi_k = s_i | w) = \frac{P(\pi_k = s_i, w)}{P(w)} =$$

анти逆行 нисходящий
следующий

forward backward
существующий P убывающий

$$= \frac{P(\pi_k = s_i, w, \dots, w_k, \dots, w_L)}{P(w)} = \frac{P(\pi_k = s_i, w, \dots, w_k) \cdot P(w_{k+1}, \dots, w_L)}{P(w)}$$

Это более простой динамического программирования (что упрощает задачу).