ФИЛЬТР БЛУМА

анализ вероятности ε
ложно-положительных срабатываний

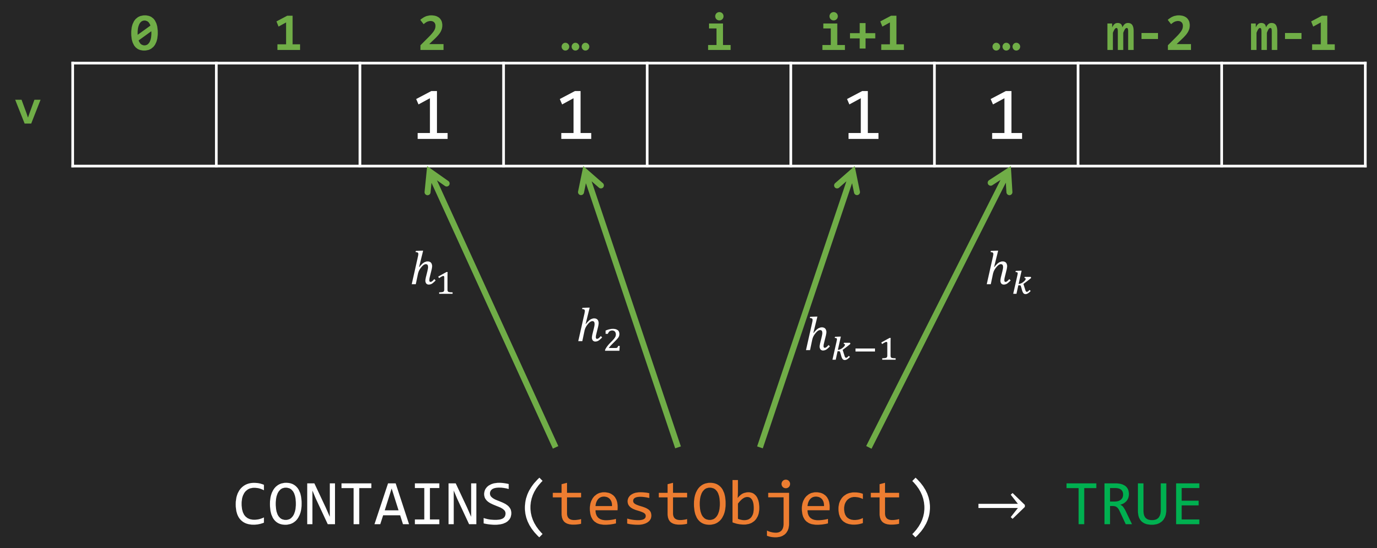
Множество объектов $S = \{s_1, s_2, s_3, \dots, s_n\}$
Битовый массив размера m

Множество хеш-функций $H = \{h_1, h_2, h_3, h_k\}$, где
$h_i : S \to \{0, 1, \dots, m-1\}$

- Хеш-функции независимы
- Хеш-функции отвечают требования равномерного распределения

Все события, совместная вероятность которых рассматривается далее, предполагаются **независимыми**

Ложно-положительное срабатывание фильтра для некоторого объекта возможно, если **по всем хеш-кодам** имеем истину

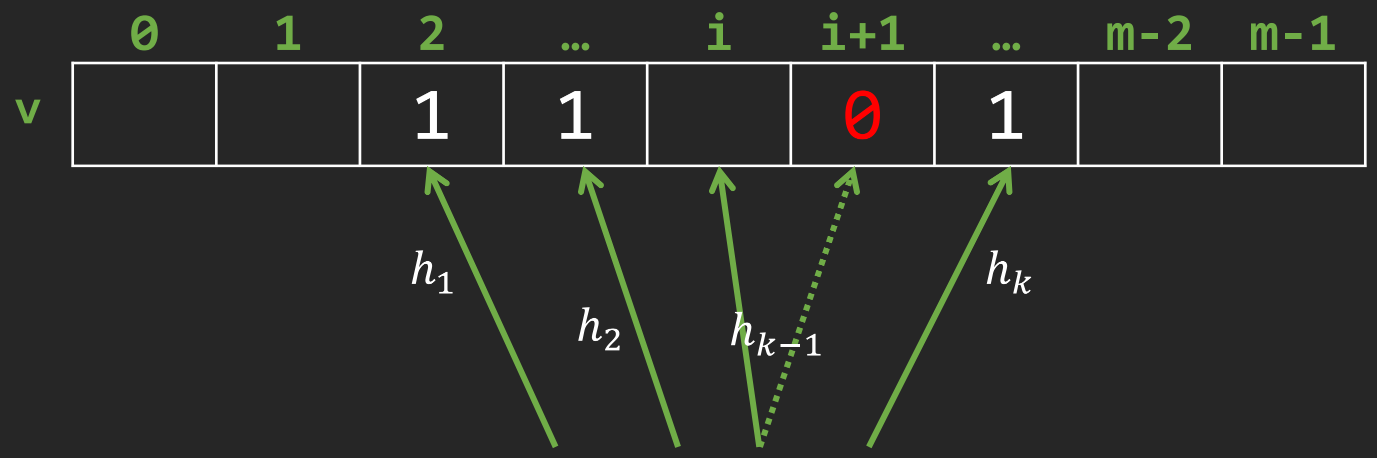


Нам достаточно **одного ложного бита**, чтобы это не являлось ложно-положительным срабатыванием

С учетом свойств k используемых хеш-функций, а также вставки n объектов в фильтр размера m

Конкретная хеш-функция НЕ попала в эту ячейку

$$P(v[i+1]=0) = \left(1 - \frac{1}{m}\right)^{k \cdot n}$$



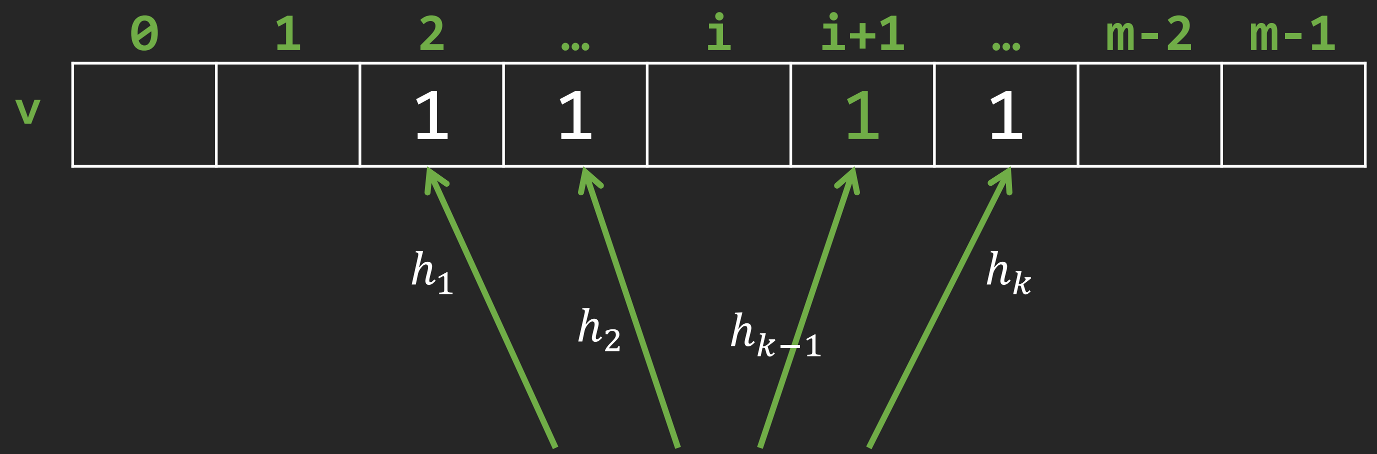
`CONTAINS(testObject)` → FALSE

Противоположное событие – значение в этом бите **истинно**

С учетом свойств k используемых хеш-функций, а также вставки n объектов в фильтр размера m

$$P(v[i+1]=1) = 1 - P(v[i+1]=0)$$

$$P(v(i+1)=1) = 1 - \left(1 - \frac{1}{m}\right)^{k \cdot n}$$



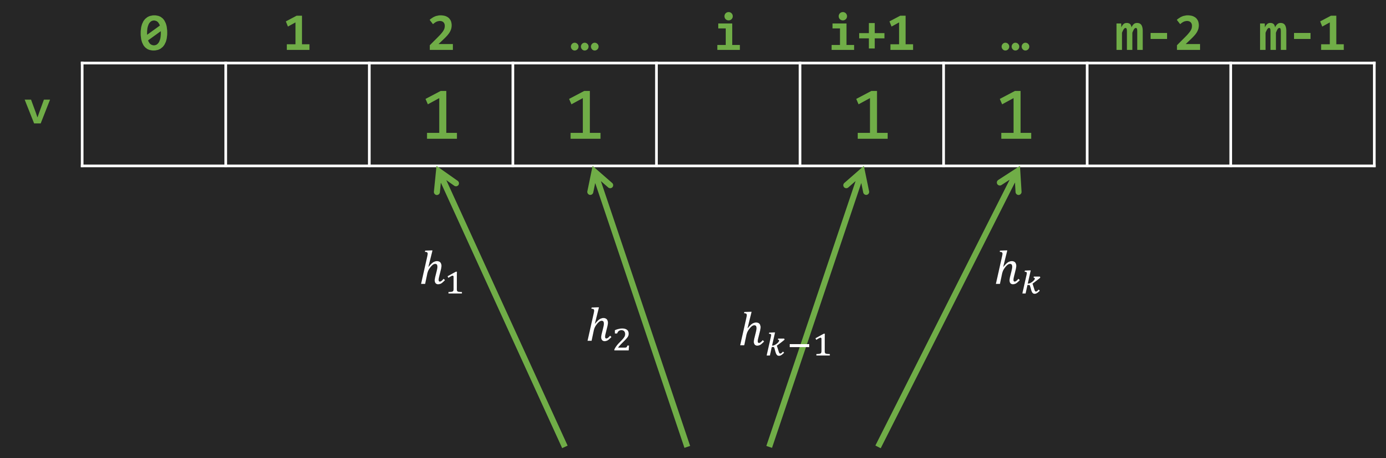
CONTAINS(testObject) → TRUE

Для ложно-положительного срабатывания значения **во всех** k битах должны быть истинными

С учетом предположения онезависимости событий

$$\varepsilon = P(v[h_1(\text{testObject})] = 1) \cdot \ldots \cdot P(v[h_k(\text{testObject})] = 1)$$

$$\varepsilon = \left(1 - \left(1 - \frac{1}{m}\right)^{k \cdot n}\right)^{k}$$



CONTAINS(testObject) → TRUE