Дискретная математика. Теория

Александр Сергеев

1 Дискретная теория вероятности

1.1 Введение

Определение
Дискретное вероятностное пространство – пара (Ω, p),
где Ω - не более чем счетное множество элементарных исходов
$p : \Omega \to [0, 1], \sum\limits_{\omega\in\Omega} p(\omega) = 1$

Определение
Событие – $A \subset \Omega$
Вероятность события – $P(A) = \sum\limits_{a\in A} p(a)$

Определение
События A и B *независимые*, если $P(A \cap B) = P(A)P(B)$

$P(B|A) = \dfrac{P(A \cap B)}{P(A)}$ - вероятность B при условии A

Лемма
Если A и B независимы, то $P(B|A) = P(B)$

Определение
Пусть $(\Omega_1, p_1), (\Omega_2, p_2)$ - независимые ДВП
Тогда их произведение $(\Omega = \Omega_1 \times \Omega_2, p(\omega_1 \in \Omega_1, \omega_2 \in \Omega_2) = p_1(\omega_1)p_2(\omega_2))$

Определение
$A_1, A_2, \dots, A_n)$ - *независимы в совокупности*,
если $\forall\, I \subset \{1 \dots n\}$ $P(\bigcap\limits_{i\in I} A_i) = \prod\limits_{i\in I} P(A_i)$

Теорема
$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$

Формула Байеса
$P(A_i|B) = \dfrac{P(B|A_i)P(A_i)}{P(B)} = \dfrac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$

Оффтоп
Рассмотрим пример: выкинули два честных кубика
Заметим, что возможно построить две математические модели:

1. Результаты - упорядоченная пара $\langle x, y\rangle$
   Тогда $|\Omega| = 36, p(w) = \dfrac{1}{36}$

2. Результаты - неупорядоченная пара $[x, y]$
   Тогда $|\Omega| = 21, p([x, x]) = \dfrac{1}{36}, p([x, y \neq x]) = \dfrac{1}{18}$

Заметим, что для запросов, не содержащих информацию об упорядоченности, результат не зависит от построенной модели

1.2 Случайные величины

Определение
Случайной величиной называется функция $\xi : \Omega \to \mathbb{R}$
Примеры случайных величин

1. Пусть кинули 10 монет. Построим случайную величину – количество выпавших орлов: $\xi(b_1, \ldots, b_n) = b_1 + \ldots + b_n$

2. Пусть кинули n кубиков. Построим случайную величину – среднее значение: $\xi(v_1, \ldots, v_n) = \dfrac{v_1 + \ldots + v_n}{n}$

3. Пусть n студентов приходят на лекцию с вероятностями $p_1, \ldots, p_n$. Построим случайную величину - количество студентов на лекции: $\xi(s_1, \ldots, s_n) = \sum_{i=1}^{n} s_i$

Заметим, что у этой случайной величины неравномерное распределение вероятностей: $p(s_1, \dots, s_n) = \prod\limits_{i=1}^{n} \begin{cases} p_i, & s_i = 1 \\ 1 - p_i, & s_i = 0 \end{cases}$

Давайте внализировать события через их случайные величины
Заметим, что уравнение $\xi = 3$ задает событие $\{w : \xi(w) = 3\}$(аналогично и другие предикаты с ξ задают события)
Определение
$f_\xi : \mathbb{R} \to \mathbb{R}, f_\xi(a) = P(\xi = a)$ – дискретная плотность распределения
$F_\xi : \mathbb{R} \to \mathbb{R}, f_\xi(a) = P(\xi \le a)$ – функция распределения
Определение
Пусть ξ – случайная величина
$E_f = \sum\limits_{\omega \in \Omega} \xi(\omega) p(\omega)$ – математическое ожидание
$E_\xi = \sum\limits_{\omega \in \Omega} \xi(\omega) p(\omega) = \sum\limits_a \sum\limits_{\omega : \xi(\omega) = a} = \sum\limits_a a \sum\limits_{\omega : \xi(\omega) = a} p(\omega) = \sum\limits_a a P(\xi = a) = \sum\limits_a a f_\xi(a)$
Определение
$D_\xi = E((\xi - E\xi)^2)$ – дисперсия
Свойства математического ожидания

1. $E(c\xi) = cE_\xi$
2. $E(\xi + \eta) = E_\xi + E_\eta$ (даже для зависимых величин)
3. Для независимых ξ, η $E(\xi\eta) = E(\xi)E(\eta)$
   Доказательство
   $E(\xi\eta) = \sum\limits_a aP(\xi\omega = a) = \sum\limits_x \sum\limits_y xyP(\xi = x \wedge \omega = y) = \sum\limits_x x \sum\limits_y yP(\xi = x \wedge \omega = y) = \sum\limits_x x \sum\limits_y yP(\xi = x)P(\omega = y) = \sum\limits_x xP(\xi = x) \sum\limits_y yP(\omega = y) = E_\xi E_\omega$
4. $E(\xi - E_\xi) = E\xi - EE\xi = E\xi - E\xi = 0$
5. $D_\xi = E((\xi - E\xi)^2) = E(\xi^2 - 2\xi E\xi + (E\xi)^2) = E(\xi^2) - E(2\xi E\xi) + E((E\xi)^2) = E(\xi^2) - (E\xi)^2$

Определение
ξ, η независимы, если $\forall\, a, b$ события $\xi = a$ и $\eta = b$ независимы
Для непрерывных величин вместо $=$ *берем* $\leq$

Пример 1
Бросаем два кубика
$\xi = v_1 + v_2$
$E_\xi = 7$
Пример 2
Бросаем кубик
$\xi = up + down$
$E_\xi = 7$
Пример 3
Ω – перестановки n элементов
$p(\sigma) = \dfrac{1}{n!}$
$\xi(\sigma) = |\{i : \sigma_i = i\}|$
Утверждается, что $E_\xi = 1$
Посчитать это через подсчет случаев сложно
Несмотря на это, мы можем посчитать матожидание
Пусть $\xi_i = (\sigma_i = i)$
$\xi = \xi_1 + \ldots + \xi_n$
$E_{\xi_i} = P(\xi_i = 1) = \dfrac{(n-1)!}{n!} = \dfrac{1}{n}$
Отсюда $\xi = n\dfrac{1}{n} = 1$

Свойства дисперсии $D(c\xi) = c^2 D(\xi)$
Дисперсия не линейна

Теорема
$D(\xi + \eta) = D(\xi) + D(\eta)$
Доказательство
$D(\xi + \eta) = E(\xi + \eta)^2 - E((\xi + \eta)^2) = E(\xi^2 + 2\xi\eta + \eta^2) - (E\xi)^2 - 2E\xi E\eta - (E\eta)^2 = E\xi^2 + 2E\xi E\eta + E\eta^2 - (E\xi)^2 - 2E\xi E\eta - (E\eta)^2 = E\xi^2 - (E\xi)^2 + E\eta^2 - (E\eta)^2 = D(\xi) + D(\eta)$
Следствие
$\xi_1, \ldots, \xi_n$ – одинаково распределенные независимые случайные величины

$\xi = \frac{1}{n}\sum_{i=1}^{n}\xi_i$

$E_\xi = E_{\xi_i}, D_\xi = \frac{1}{n}D_{\xi_i}$

Определение

$\sigma = \sqrt{D_\xi}$ – среднеквадратичное отклонение

1.3 Хвостовые неравенства

Неравенство Маркова

Пусть $\xi \geq 0, E_\xi > 0$

Оценим $P(\xi \geq cE_\xi) \leq \frac{1}{c}$

Доказательство

$P(\xi \geq cE_\xi) = \sum_{\substack{\omega \\ \xi(\omega) \geq cE_\xi}} p(\omega)$

$E_\xi = \sum_{\omega} p(\omega)\xi(\omega) = \sum_{\substack{\omega \\ \xi(\omega) \geq cE_\xi}} p(\omega)\xi(\omega) + \sum_{\substack{\omega \\ \xi(\omega) < cE_\xi}} p(\omega)\xi(\omega) \geq \sum_{\substack{\omega \\ \xi(\omega) \geq cE_\xi}} p(\omega)\xi(\omega) \geq$

$cE_\xi \sum_{\substack{\omega \\ \xi(\omega) \geq cE_\xi}} p(\omega) = cE_\xi P(\xi \geq cE_\xi)$

$1 \geq cP(\xi \geq cE_\xi)$

$P(\xi \geq cE_\xi) \leq \frac{1}{c}$

Неравенство Чебышева

$P(|\xi - E_\xi| \geq c\sqrt{D_\xi}) \leq \frac{1}{c^2}$ – относительная форма неравенства Чебышева

Доказательство

Возьмем $\eta = (\xi - E_\xi)^2$

Неравентсво Чебышева (ver. 2)

$c := \frac{a}{\sqrt{D_\xi}}$

$P(|\xi - E_\xi| \geq a) \leq \frac{D_\xi}{a^2}$ – абсолютная форма неравенства Чебышева

Пример

Возьмем честную монету

$E_\xi = \frac{1}{2}$
$D_\xi = \frac{1}{4}$
$D_\xi = E_\xi - (E_\xi)^2$
$P(|\xi - E_\xi| \geq \frac{1}{2}) \leq 1$
$P(|\xi - E_\xi| \geq 1) \leq \frac{1}{4}$ (на самом деле 0)
Видим, что оценка сверху неточная

Пример 2
$\xi_1, \ldots, \xi_n$ – одинаково распределенные независимые случайные величины
$\xi = \frac{1}{n}\sum_{i=1}^{n}\xi_i$
$P(|\xi - E_\xi| \geq \varepsilon) \leq \frac{D_\xi}{\varepsilon^2}$
$P(|\xi - E_\xi| \geq \varepsilon) \leq \frac{D_{\xi_i}}{n\varepsilon^2}$
Пусть мы хотим не попадать в ε-окрестность с вероятностью не более δ (вероятность промаха)
$P(|\xi - E_\xi| < \varepsilon) > 1 - \delta$
$P(|\xi - E_\xi| \geq \varepsilon) \leq \delta$
Тогда $\frac{D_{\xi_i}}{n\varepsilon^2} \leq \delta$
$n \geq \frac{D_{\xi_i}}{\varepsilon^2\delta} \sim \frac{1}{\varepsilon^2\delta}$
Граница Чернова для монеты Бернулли
$P(\xi \geq (1+\delta)np) \leq e^{-\frac{\delta^2}{2+\delta}np}$
$P(\xi \leq (1-\delta)np) \leq e^{-\frac{\delta^2}{2}np}$
Доказательство
Доказательства не будет, жди теорвер

1.4 Введение в информатику

Определение
информация = -неопределенность (по Шеннону)

Рассмотрим модель случайного источника
Пусть есть вероятностное пространство Ω и распределение p
Получая событие ω, мы получаем информацию о том, что оно произошло
Определим, сколько информации мы получаем
Заметим, что оно не зависит от Ω
Пусть $H(p_1, p_2, \ldots)$ – количество информации в зависимости от вероятностей событий
H удовлетворяет следующим свойствам:

1. Для любого числа n $H(p_1, \ldots, p_n)$ – непрерывно
   (т.к. при малом изменении вероятностей количество информации мало изменяется)

2. $H(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}) = h(n)$
   $h(n) \uparrow$ – (т.к. чем больше вариантов, тем больше информации)

3. Аддитивность
   Пусть $\Omega \subset \mathbb{R}^2$ – множество пар
   $A_a = \{(a, *) \in \Omega\}$
   $P(A_a) = p_a$
   $P(\{(a,b)\}|A_a) = q_{a,b}$
   $p(\{(a,b)\}) = p_a q_{a,b}$
   Тогда $H(p_1q_{1,1}, \ldots, p_1q_{1,k_1}, \ldots, p_nq_{n,1}, \ldots, p_nq_{n,k_n}) = H(p_1, \ldots, p_n) + \sum\limits_i p_i H(q_{i,1}, \ldots, q_{i,k_i})$

Рассмотрим случай с равными вероятностями
$h(nm) = h(n) + \sum\limits_{i=1}^{n} \frac{1}{n} h(m) = h(n) + h(m)$
Лемма
$h(n) = c\log_2(n)$
Традиционно $c = h(2)$ – бит
Доказательство
$h(2^k) = kh(2) = ck$
Рассмотрим $n^t, n, t \in \mathbb{N}$
Пусть $2^k \leq n^t < 2^{k+1}$

Тогда $h(2^k) \leq h(n^t) \leq h(2^{k+1})$
$ck \leq th(n) \leq c(k+1)$
$\frac{ck}{t} \leq h(n) \leq \frac{c(k+1)}{t}$
$k \leq t\log_2(n) \leq k+1$
$\frac{ck}{t} \leq c\log_2(n) \leq \frac{c(k+1)}{t}$
$|h(n) - c\log_2(n)| \leq \frac{c}{t}$ – при всех t
Отсюда $h(n) = c\log_2(n)$

"Разберемся"с H
Начнем с $p_i, q_i \in \mathbb{Q}$
Пусть $p_i = \frac{a_i}{b}$
$k_i = a_i, q_{ij} = \frac{1}{a_i}$
Отсюда $H(\frac{1}{b}, \ldots, \frac{1}{b}) = H(p_1, \ldots, p_n) + \sum\limits_{i=1}^{n} p_i H(\frac{1}{a_i}, \ldots, \frac{1}{a_i}) = H(p_1, \ldots, p_n) + \sum\limits_{i=1}^{n} c\log_2(a_i)$
$H(p_1, \ldots, p_n) = -c(\sum\limits_{i=1}^{n} p_i \log_2(a_i) - \log_2(b)) = -c(\sum\limits_{i=1}^{n} p_i \log_2(a_i) - \sum\limits_{i=1}^{n} p_i \log_2(b)) = -c\sum\limits_{i=1}^{n} p_i(\log_2(a_i) - \log_2(b)) = -c\sum\limits_{i=1}^{n} p_i \log_2(\frac{a_i}{b}) = -c\sum\limits_{i=1}^{n} p_i \log_2(p_i)$
Из непрерывности формула верна для всех $p_i \in \mathbb{R}$
Выберем $c = 1$ бит
Тогда $H = -\sum\limits_{i=1}^{n} p_i \log_2(p_i)$ бит
Или $H = \sum\limits_{i=1}^{n} p_i \log_2(\frac{1}{p_i})$ бит – энтропия
Флешбеш: арифметическое кодирование использует в среднем H бит на каждый символ
Отсюда арифметическое кодирование - оптимальное кодирование для данных, которые можно аппроксимировать случайным источником
Ограничение в H бит на символ называют *энтропийным барьером*
Энтропийный барьер можно преодолеть лишь учетом закономерностей

в последовательности символов

Энтропия Шеннона хорошо описывает случайные последовательности и плохо описывает "регулярные"строки (строчки, имеющие закономерности)
Для измерения информации в более сложных объектах используется *Колмогоровская сложность*
Колмогоровская сложность зависит от *декодера* и равна количеству информации, необходимому для кодирования объекта
$K_A(s) \leq K_B(s) + C_{A,B}$, где A, B – декодеры, C – константа
$K(s) \leq H(s) + C$

2 Цепи Маркова

2.1 Введение

Определение
Марковская цепь – взвешенный ориентированный граф с неотрицательными весами и суммарным весом исходящих ребер, равным 1
Пронумеруем состояния (вершины)
Пусть $b = \begin{pmatrix} b_1 & b_2 & \dots & b_n \end{pmatrix}$ – матрица состояния B, где b_i – вероятность находиться в i-ом состоянии ($P(B = i)$)
Пусть $c = \begin{pmatrix} c_1 & c_2 & \dots & c_n \end{pmatrix}$ – матрица состояния C
Рассмотрим матрицу переходов $P = (p_{ij})_{n \times n}$, где p_{ij} – вероятность перейти из i в j
Найдем зависимость между b и c
$c_i = P(C = i) = \sum\limits_{j=1}^{n} P(C = i | B = j) P(B = j) = \sum\limits_{j=1}^{n} p_{ji} b_j$
Отсюда $c = b \cdot P$
Тогда распределение вероятностей на i-ом шаге $b^i = b^0 P^i$, где b^0 – начальное состояние

Рассмотрим цепь Маркова как граф
Вершина в цепи Маркова называется *состоянием*
Поглощающее(существенное) состояние – состояние с кольцевым ребром веса 1

Цепь Маркова называется *поглощающей*, если из любого состояния можно попасть в поглощающее
Пример непоглощающей цепи: цепь с циклом длины 2 и более, где все ребра веса 1
Эргодический класс – компонента сильной связности графа Марковских цепей
Компонента сильной связности – максимальное по включению множество вершин, где из каждой можно дойти до каждой
(класс эквивалентности для отношения достижимости)
Эргодический класс называется *поглощающим*, если из него нет исходящих переходов
Цепь Маркова можно представить как граф эргодических классов (но оценить веса ребер не всегда просто)
Цепь Маркова называется *эргодической*, если она содержит ровно 1 эргодический класс

Эргодический класс называется *периодическим с циклом* d, если любая длина цикла в этом классе делится на $d > 1$. Иначе – *регулярной*

Теорема о классификации Марковских цепей
Любая Марковская цепь содержит поглощающие эргодические классы
Марковская цепь с вероятностью 1 рано или поздно оказывается в состоянии из поглощающего эргодическим классом
Для непериодического поглощающего эргодического класса в случае попадания в него существует предельное распределение вероятностей b : $b = bP$
Для любого распределения b^0 $b^0 P^n \to b$

Для цепей Маркова существуют две независимые задачи: задача поглощения – задача определения, в какой поглощающий эргодический класс мы попадем, и задача стационарного распределения внутри поглощающего эргодического класса
Займемся задачей поглощения (т.е. определим, в какой эргодический класс мы попадем)
В ходе решения этой задачи поглощающие эргодический классы можно заменить на одно поглощающее состояние
Занумеруем состояния так, чтобы сначала шли непоглощающие, а потом поглощающие

Пусть $1 \dots m$ – непоглощающие состояния, $m+1 \dots n$ – поглощающие
Тогда $P = \begin{pmatrix} Q & R \\ \mathbb{0} & I \end{pmatrix}$, где
$Q = P[1 \dots m][1 \dots m]$
$R = P[1 \dots m][m+1 \dots n]$
$\mathbb{0} = P[m+1 \dots n][1 \dots m]$ – нулевая матрица
$I = P[m+1 \dots n][m+1 \dots n]$ – единичная матрица
Возьмем матрицу состояния $b = \begin{pmatrix} b_1 & b_2 & \dots & b_n \end{pmatrix}$
Пусть $a = \begin{pmatrix} b_1 & b_2 & \dots & b_m \end{pmatrix}$
$a^n = a^0 Q^n$
Лемма
$Q^n \to \mathbb{0}$
Доказательство
Пусть L – максимальная длина кратчайшего пути от i до поглощающей
Найдем $X = Q^L$
$x_{ij} = \sum\limits_{k_1, k_2, \dots, k_{L-1}} q_{ik_1} q_{k_1 k_2} \dots q_{k_{L-1} j}$
$\sum\limits_{j\text{– непогл.}} x_{ij} = \sum\limits_{k_1, k_2, \dots, k_{L-1}, j} q_{ik_1} q_{k_1 k_2} \dots q_{k_{L-1} j} = \delta_i < 1$ – т.к. это вероятность пройти от i до непоглощающего состояния (если бы до любого состояния, то было бы 1)
Отсюда $\max\limits_{i=1\dots m} \sum\limits_{j\text{– непогл.}} x_{ij} = \max \delta_i = \delta < 1$
Тогда $Q^n = Q^L Q^{n-L}$
Пусть $\max Q^{n-L} = v_{n-L}$
$Q^n_{ij} = (Q^L Q^{n-L})_{ij} = \sum\limits_k Q^L_{ik} Q^{n-L}_{kj} \le \sum\limits_k Q^L_{ik} v_{n-L} \le \delta v_{n-L}$
$v_n \le \delta^{\lfloor \frac{n}{L} \rfloor} \to 0$
Тогда $Q^n \to \mathbb{0}$
Теорема о поглощении
Поглощающая Марковская цепь переходит в состояние поглощения с вероятностью 1
Доказательство
Следует из леммы

Научимся определять, где же мы поглотимся
Для этого найдем мат. ожидание времени до поглощения

b^0 – начальное распределение
T – случайная величина – число шагов до поглощения
$T = \sum\limits_{i=1}^{m} T_i$, где T_i – число посещений i-ого состояния
$T_i = \sum\limits_{j=0}^{\infty} T_{ij}$, где $T_{ij} = \begin{cases} 1, & \text{если на } j\text{-ом шаге мы в состоянии } i \\ 0, & \text{иначе} \end{cases}$

Лемма

$$\sum_{j=0}^{\infty} Q^j = (I - Q)^{-1}$$

Доказательство

$(I-Q)(I+Q+Q^2+\ldots+Q^n) = I+Q+Q^2+\ldots+Q^n-Q-Q^2-\ldots-Q^{n+1} = I - Q^{n+1} \to I$

Определение

$N = (I-Q)^{-1}$ – фундаментальная матрица поглощения Марковской цепи

$$ET = \sum_{i=1}^{m} ET_i = \sum_{i=1}^{m}\sum_{j=0}^{\infty} ET_{ij} = \sum_{i=1}^{m}\sum_{j=0}^{\infty} P(\text{цепь в состоянии } i \text{ на шаге } j) =$$
$$\sum_{i=1}^{m}\sum_{j=0}^{\infty}(a^0Q^j)_i = \sum_{i=1}^{m}(\sum_{j=0}^{\infty} a^0Q^j)_i = \sum_{i=1}^{m}(a^0\sum_{j=0}^{\infty} Q^j)_i = \sum_{i=1}^{m}(a^0N)_i = a^0N\mathbb{1}$$

Заметим, что $a^0N = \begin{pmatrix} ET_1 & ET_2 & \ldots & ET_m \end{pmatrix}$

$$P(\text{погл. в } j) = \sum_{i=1}^{m} P(\text{погл. в } j \text{ из } i)P(\text{быть в } j) = \sum_{t=0}^{\infty}\sum_{i=1}^{m} P(\text{погл. в } j \text{ из } i)P(\text{быть в } i \text{ на шаге}$$
$$\sum_{t=0}^{\infty}\sum_{i=1}^{m} R_{i,j-m}P(\text{быть в } i \text{ на шаге } t) = \sum_{i=1}^{m} R_{i,j-m}\sum_{t=0}^{\infty} P(\text{быть в } i \text{ на шаге } t) =$$
$$\sum_{i=1}^{m}(a^0N)_iR_{i,j-m} = (a^0NR)_{j-m}$$

Отсюда $A = \begin{pmatrix} P(\text{погл. в } m+1) & P(\text{погл. в } m+2) & \ldots & P(\text{погл. в } n) \end{pmatrix} = a^0NR$

Эргодическая теорема для регулярных цепей

Пусть Марковская цепь такова, что $\forall\, i, j\ p_{ij} > 0$ (данная цепь непериодическая)
Тогда $\exists\, b\ \forall\, b^0\ b^0P^n \to b$
(Отсюда $b = bP$, т.к. $bP = \lim\limits_n bP^{n+1} = \lim\limits_n bP^n = b$)

Доказательство
Рассмотрим $b^0 A$:
Предположим, что $\forall\, j\ a_{ji} = \widetilde{a}_i$

$$(b^0 \cdot A)_i = \sum_{j=1}^{n} b^0_j a_{ji} = \underbrace{\sum_{j=1}^{n} b^0_j}_{1} \widetilde{a}_i = \widetilde{a}_i$$

Докажем, что $P^t \to A : \forall\, j\ a_{ji} = \widetilde{a}_i$
Пусть $m_i^n = \min\limits_j (P^t)_{ji}$
$M_i^n = \max\limits_j (P^t)_{ji}$
Лемма
$M_i^t - m_i^t \to 0$
Доказательство
$\delta := \min\limits_{ij} p_{ij}, \delta > 0$(из условия теоремы)
Рассмотрим P^{t+1}:

$$p_{ji}^{t+1} = \sum_{k=1}^{n} p_{jk} p_{ki}^t \leq \sum_{k=1, k\neq \mathrm{posMin}}^{n} p_{jk} M_i^t + p_{j\,\mathrm{posMin}} m_i^t = \underbrace{\sum_{k=1}^{n} p_{jk}}_{1} M_i^t + p_{j\,\mathrm{posMin}} (m_i^t - M_i^t) \leq M_i^t + \delta(m_i^t - M_i^t)$$

Аналогично $m_i^t + \delta(M_i^t - m_i^t) \leq p_{ji}^{t+1}$
Отсюда $M_i^{t+1} - m_i^{t+1} \leq (M_i^t - m_i^t)(1 - 2\delta) \leq (1 - 2\delta)^{t+1} \to 0$

Научимся искать b
$bP = b$
Заметим, что у данном системы есть одно или бесконечно много решений
Утверждается, что $\operatorname{rg} I - P = n - 1$
Тогда пространство решений одномерное
Тогда $\exists\, ! b : \sum\limits_i b = 1$

Т.о. найти b можно двумя способами:

1. $b = \lim\limits_n P^n$
2. $b : (I - P) b = \mathbb{0}, \sum b_i = 1$ – СЛОУ

Соединим теоремы:
Пусть у нас есть Марковская цепь без периодических классов

Для начала представим, что внутри всех поглощающих классов сами состояния являются поглощающими(т.е. удалим внутренние ребра поглощающих классов и добавим петли)
Теперь мы можем определить вероятность попадания в каждое состояние каждого поглощающего класса
b^0NR – наше распределение
Теперь рассмотрим эргодический класс A
Пусть $\widetilde{p} = \sum\limits_{a \in A} (b^0NR)_a$
$\widetilde{b}^0 = \sum\limits_{a \in A} (b^0NR)_a \Big|_A \dfrac{1}{\widetilde{p}}$ – начальное состояние внутри эргодического класса A
По теореме $\exists\, b : \widetilde{b}^0 A^n \to b$
Тогда конечное распределение – объединение всех $b\widetilde{p}$