КРИТЕРИЙ ЛЕМПЕЛЯ–ЗИВА — НОВАЯ НАДЕЖДА

Костевич Андрей Леонидович, канд. физ.-мат наук,
руководитель группы разработки средств защиты информации ЗАО “Авест”

Шилкин Антон Владимирович, канд. физ.-мат наук,
зав. сектором компьютерной безопасности НИИ прикладных проблем математики и информатики

Москва, 2013

I. Введение

$\mathcal{H}_0 : \{X_t\}$ — н.о.р. с.в. Бернулли, $\mathbf{P}\{X_t = 1\} = \frac{1}{2}$

Актуальность задачи в криптографии:

- предварительный анализ стойкости криптографических алгоритмов: тестирование выходных последовательностей
  AES: 4 из 15 (25%) кандидатов не прошли тестирование
  NESSIE: 13 из 31 (42%) кандидатов не прошли тестирование
- проверка работоспособности физических ДСЧ: тестирование ключевых последовательностей
  - AIS 31 — стандарт ФРГ с требованиями к физическим ДСЧ
  - FIPS 140-2 — стандарты США с требованиями к СКЗИ
- оценка энтропии физических ДСЧ
  - NIST SP800-90 — стандарт США с требованиями к энтропии

I. Введение

$\mathcal{H}_0 : \{X_t\}$ — н.о.р. с.в. Бернулли, $\mathbf{P}\{X_t = 1\} = \frac{1}{2}$

Традиционные подходы:

- Эмпирический (Д. Кнут, G. Marsaglia): предлагается статистика, находится ее распределение при верной $\mathcal{H}_0$
  Особенность: выявляемая альтернатива неизвестна
- Параметрический (Г. И. Ивченко, Ю. И. Медведев, В. П. Чистяков, А. М. Зубков): выбирается параметрическое семейство альтернатив, оптимальным образом строится критерий
  Особенность: на практике требуется рассматривать большое число параметрических семейств

Сложностной подход: Y. Ziv, A. Lempel, U. Maurer, Б. Рябко
Использование мер сложности последовательности: случайная посл. не может быть сжата или она является сложной для описания.
Достоинство: $\mathcal{H}_1$ включает большое число вероятностных моделей
Особенность: сложность исследования вероятностных характеристик

Основные наборы статистических тестов

Автор	Год издания	Критериев	Сложностных
FIPS 140-1	1994	4	-
NESSIE	2000	20	2
NIST SP 800-22	2001(2010)	16(15)	2(1)
AIS 31	2001	9	1

Реализации критериев в рамках сложностного подхода:

- Критерий сжатия Лемпеля-Зива (NESSIE, NIST SP 800-22):
  основан на универсальном алгоритме сжатия Лемпеля-Зива
  Критерий Маурера (NESSIE, NIST SP 800-22, AIS 31):
  основан на универсальном алгоритме сжатия Elias-Willems
  Проблема с теоретическим обоснованием: вероятностные характеристики оценены с исп. стат. методов
- Критерий Б.Я. Рябко:
  основан на любом универсальном алгоритме сжатия
  Проблема в сложности исследования вероятностных характеристик

Обоснование выбора предикторов

Общая проблема — сложность записи универсальных алгоритмов сжатия приводит к сложности исследования вероятностных свойств статистик критериев, поэтому предлагается использовать универсальные предикторы

Аргументы выбора подхода на базе предикторов:

- теорема A. Yao
  в рамках теории сложности доказана эквивалентность всех имеющих полиномиальную сложностью статистических критериев проверки случайности одному критерию: вероятность успеха прогноза следующего бита (next-bit test) отлична от 0.5
- связь с мерами сложности: по алгоритму сжатия информации может быть построен предиктор и наоборот
- инструмент оценки энтропии

Построение критерия и его основные свойства

Наблюдается $X_1^\infty = X_1, X_2, \ldots$, где $X_t \in \mathcal{A} = \{0, 1\}$, описывается условными вероятностями $\{\mathbf{P}\{X_{t+1} \mid X_1^t\}\}$ из класса моделей $\mathcal{M}$

Если вероятностная модель известна, то используется МП-предиктор:

$$\hat{X}_{t+1}^* = \arg\max_{a \in \mathcal{A}} \mathbf{P}\left\{X_{t+1} = a \mid X_1^t\right\}, \quad \underset{\text{вер.ош.}}{\pi_t^*(X_1^t)} = \min_{a \in \mathcal{A}} \mathbf{P}\left\{X_{t+1} = a \mid X_1^t\right\} \tag{1}$$

Если вероятностная модель неизвестна, то использование предиктора позволяет построить оценки $\{\hat{\mathbf{P}}\{X_{t+1} \mid X_1^t\}\}$ и:

$$\hat{X}_{t+1} = \arg\max_{a \in \mathcal{A}} \hat{\mathbf{P}}\{X_{t+1} = a \mid X_1^t\}, \quad \underset{\text{вер.ош.}}{\hat{\pi}_t(X_1^t) \geqslant \pi_t^*(X_1^t)} \tag{2}$$

Предиктор (2) называется универсальным для класса $\mathcal{M}$, если при неизвестных $\{\mathbf{P}\{X_{t+1} \mid X_1^t\}\}$ из класса $\mathcal{M}$ **ошибка предсказания стремится к нулю**:

$$\hat{\pi}_t(X_1^t) - \pi_t^*(X_1^t) \xrightarrow{\mathbf{P}} 0, t \to \infty \tag{3}$$

Построение критерия и его основные свойства

Предиктор (2) будем называть состоятельным для класса $\mathcal{M}$, если

$$\arg\max_{a\in\mathcal{A}} \hat{\mathbf{P}}\{X_{t+1} = a \mid X_1^t\} - \arg\max_{a\in\mathcal{A}} \mathbf{P}\left\{X_{t+1} = a \mid X_1^t\right\} \xrightarrow{\mathbf{P}} 0, \quad t \to \infty$$

По мере регистрации X_t с помощью предиктора вычисляются прогнозы $\hat{X}_{t+1}$ и индикаторы успехов прогноза $Y_t = \mathbf{I}\left\{\hat{X}_t = X_t\right\}$:

$$\begin{array}{lll} X_0, & X_1, & \\ & | & \\ & Y_1 = & \mathbf{I}\{\hat{X}_1 \quad = X_1\} \end{array}$$

Построение критерия и его основные свойства

Предиктор (2) будем называть состоятельным для класса $\mathcal{M}$, если

$$\arg\max_{a\in\mathcal{A}} \hat{\mathbf{P}}\{X_{t+1}=a \mid X_1^t\} - \arg\max_{a\in\mathcal{A}} \mathbf{P}\left\{X_{t+1}=a \mid X_1^t\right\} \xrightarrow{\mathbf{P}} 0, \quad t\to\infty$$

По мере регистрации X_t с помощью предиктора вычисляются прогнозы $\hat{X}_{t+1}$ и индикаторы успехов прогноза $Y_t = \mathbf{I}\left\{\hat{X}_t = X_t\right\}$:

$$\begin{array}{ccccc} X_0, & X_1, & & X_2, & \\ & \Big| & & \Big| & \\ & Y_1 & & Y_2 = & \mathbf{I}\{\hat{X}_2 = X_2\} \end{array}$$

Построение критерия и его основные свойства

Предиктор (2) будем называть состоятельным для класса $\mathcal{M}$, если

$$\arg\max_{a\in\mathcal{A}} \hat{\mathbf{P}}\{X_{t+1}=a \mid X_1^t\} - \arg\max_{a\in\mathcal{A}} \mathbf{P}\left\{X_{t+1}=a \mid X_1^t\right\} \xrightarrow{\mathbf{P}} 0, \quad t\to\infty$$

По мере регистрации X_t с помощью предиктора вычисляются прогнозы $\hat{X}_{t+1}$ и индикаторы успехов прогноза $Y_t = \mathbf{I}\left\{\hat{X}_t = X_t\right\}$:

$$\begin{array}{ccccccc} X_0, & X_1, & X_2, & X_3, & & \dots \\ & \big| & \big| & \big| & & \\ & Y_1 & Y_2 & Y_3 = & \mathbf{I}\{\hat{X}_3 = X_3\} & & \end{array}$$

Построение критерия и его основные свойства

Предиктор (2) будем называть состоятельным для класса $\mathcal{M}$, если

$$\arg\max_{a\in\mathcal{A}} \hat{\mathbf{P}}\{X_{t+1}=a \mid X_1^t\} - \arg\max_{a\in\mathcal{A}} \mathbf{P}\left\{X_{t+1}=a \mid X_1^t\right\} \xrightarrow{\mathbf{P}} 0, \quad t\to\infty$$

По мере регистрации X_t с помощью предиктора вычисляются прогнозы $\hat{X}_{t+1}$ и индикаторы успехов прогноза $Y_t = \mathbf{I}\left\{\hat{X}_t = X_t\right\}$:

$$X_0, X_1, X_2, \ldots, X_t \quad \mapsto \quad Y_1^t = Y_1, Y_2, \ldots, Y_t$$

Свойство (сохранение энтропии). $H\{Y_1^t\} = H\{X_1^t\}$

Рассмотрим альтернативу $\mathcal{H}_1$ — имеются реализуемые предыстории с неравновероятными переходами:

$$\max_{a\in\mathcal{A}} \mathbf{P}\left\{X_t=a \mid X_1=x_1,\ldots,X_{t-1}=x_{t-1}\right\} = \frac{1}{2} + \varepsilon_{x_1^{t-1}}, \quad \varepsilon_{x_1^{t-1}} \geqslant 0, \tag{4}$$

$$\exists\{t_j\}, \quad \exists x_1^{*t_j} = x_1^*, \cdots, x_{t_j}^*, \text{ что } \varepsilon_{x_1^{*t_j}} > 0$$

Построение критерия и его основные свойства

Предиктор (2) будем называть состоятельным для класса $\mathcal{M}$, если

$$\arg\max_{a\in\mathcal{A}} \hat{\mathbf{P}}\{X_{t+1}=a \mid X_1^t\} - \arg\max_{a\in\mathcal{A}} \mathbf{P}\left\{X_{t+1}=a \mid X_1^t\right\} \xrightarrow{\mathbf{P}} 0, \quad t\to\infty$$

По мере регистрации X_t с помощью предиктора вычисляются прогнозы $\hat{X}_{t+1}$ и индикаторы успехов прогноза $Y_t = \mathbf{I}\left\{\hat{X}_t = X_t\right\}$:

$$X_0, X_1, X_2, \ldots, X_t \quad \mapsto \quad Y_1^t = Y_1, Y_2, \ldots, Y_t$$

Свойство (сохранение энтропии). $H\{Y_1^t\} = H\{X_1^t\}$

Рассмотрим альтернативу $\mathcal{H}_1$ — имеются реализуемые предыстории с неравновероятными переходами:

$$\max_{a\in\mathcal{A}} \mathbf{P}\left\{X_t=a \mid X_1=x_1,\ldots,X_{t-1}=x_{t-1}\right\} = \frac{1}{2} + \varepsilon_{x_1^{t-1}}, \quad \varepsilon_{x_1^{t-1}} \geqslant 0, \tag{4}$$

$$\exists\{t_j\}, \quad \exists x_1^{*t_j} = x_1^*, \cdots, x_{t_j}^*, \text{ что } \varepsilon_{x_1^{*t_j}} > 0$$

Построение критерия и его основные свойства

Предиктор (2) будем называть состоятельным для класса $\mathcal{M}$, если

$$\arg\max_{a\in\mathcal{A}}\hat{\mathbf{P}}\{X_{t+1}=a\mid X_1^t\}-\arg\max_{a\in\mathcal{A}}\mathbf{P}\left\{X_{t+1}=a\mid X_1^t\right\}\xrightarrow{\mathbf{P}}0,\quad t\to\infty$$

По мере регистрации X_t с помощью предиктора вычисляются прогнозы $\hat{X}_{t+1}$ и индикаторы успехов прогноза $Y_t=\mathbf{I}\left\{\hat{X}_t=X_t\right\}$:

$$X_0,X_1,X_2,\ldots,X_t\quad\mapsto\quad Y_1^t=Y_1,Y_2,\ldots,Y_t$$

Свойство (сохранение энтропии). $H\{Y_1^t\}=H\{X_1^t\}$

Рассмотрим альтернативу $\mathcal{H}_1$ — имеются реализуемые предыстории с неравновероятными переходами:

$$\max_{a\in\mathcal{A}}\mathbf{P}\left\{X_t=a\mid X_1=x_1,\ldots,X_{t-1}=x_{t-1}\right\}=\frac{1}{2}+\varepsilon_{x_1^{t-1}},\quad\varepsilon_{x_1^{t-1}}\geqslant 0, \tag{4}$$

$$\exists\{t_j\},\quad\exists x_1^{*t_j}=x_1^*,\cdots,x_{t_j}^*,\ \text{что}\ \varepsilon_{x_1^{*t_j}}>0$$

Построение критерия и его основные свойства

Теорема 1 (вероятностные свойства Y_t). Если предиктор является состоятельным для $\mathcal{H}_1$ (4), то случайная величина Y_t обладает следующим маргинальным распределением:

$$\begin{aligned}
\mathbf{P}\left\{Y_t = 1\right\} &= \frac{1}{2} + \hat{\varepsilon}_t, \text{ причем } \hat{\varepsilon}_t > 0 \text{ при } t > t^* \qquad (5)\\
\hat{\varepsilon}_t &= \varepsilon_t^* - (\hat{\pi}_{t-1} - \pi_{t-1}^*),\\
\varepsilon_t^* &= \sum_{i_1,\ldots,i_{t-1} \in \mathcal{A}^{t-1}} \varepsilon_{i_1^{t-1}} \cdot \mathbf{P}\left\{X_1^{t-1} = i_1^{t-1}\right\},\\
\hat{\pi}_{t-1} &= \mathbf{E}\left\{\hat{\pi}_{t-1}(X_1^{t-1})\right\}, \qquad \pi_{t-1}^* = \mathbf{E}\left\{\pi_{t-1}^*(X_1^{t-1})\right\}.
\end{aligned}$$

Пусть по последовательности X_0^n построена Y_1^n. Статистика:

$$S_n = \frac{1}{n}\sum_{t=1}^{n} Y_t, \qquad \begin{aligned} \mathcal{H}_0: &\quad \mu = \mu_0 = \frac{1}{2}, & \sigma^2 = \sigma_0^2 = \frac{1}{4n}\\ \mathcal{H}_1: &\quad \mu = \mu_1 > \frac{1}{2}, & \sigma^2 = \sigma_1^2 = \mathbf{D}\left\{S\right\} \end{aligned}$$

Построение критерия и его основные свойства

Критерий:

$$\text{принимается} \begin{cases} \mathcal{H}_0, & 2\sqrt{n}(S_n - \frac{1}{2}) < \Phi^{-1}(1-\alpha) \\ \mathcal{H}_1, & \text{иначе} \end{cases} \tag{6}$$

Теорема 2 (несмещенность). При использовании любого предиктора в асимптотике $n \to \infty$ критерий (6) имеет уровень значимости α

Теорема 3 (условие состоятельности). Пусть предиктор является состоятельным для $\mathcal{H}_1$, для $\{Y_t\}$ выполняется ЦПТ и $\mathbf{D}\{S_n\} = \sigma_n^2$ порядка $\frac{1}{n}$. Тогда при $n \to \infty$ критерий (6) является состоятельным, мощность имеет вид:

$$\left| W_n - \left(1 - \Phi\left(\frac{\Delta}{2\sqrt{n}\sigma_n} - \frac{1}{n\sigma_n}\left(\sum_{t=1}^{n} \hat{\varepsilon}_t \right) \right) \right) \right| \to 0,$$

$$W_n \to 1 \text{ при } \frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n} \hat{\varepsilon}_t \right) \xrightarrow[n\to\infty]{} \infty,$$

где $\hat{\varepsilon}_t$ определена в теореме 1.

Вероятностные свойства предиктора максимального правдоподобия для модели испытаний Бернулли

$$\mathcal{H}_1 : \mathbf{P}\{X_t = 1\} = p, \mathbf{P}\{X_t = 0\} = 1 - p, \quad p = 0.5 + \varepsilon, \quad 0 < |\varepsilon| < 0.5. \tag{7}$$

Пусть регистрируется последовательность $X_1^n = X_{-m+1}^0 || X_1^k$ объема n

По первой части X_{-m+1}^0 будем строить ОМП-оценку параметра p

$$\hat{p}_m = m^{-1} \sum_{i=-m+1}^{0} \mathbf{I}\{X_i = 1\}$$

По второй части X_1^k строится Y_1^k, где $\hat{X}_t = \begin{cases} 0 & \text{если} \quad \hat{p}_m < 0.5, \\ 1 & \text{иначе.} \end{cases}$

Справедливо представление $Y_t = X_t \oplus f(\hat{p}) \oplus 1$ — с.в. Бернулли

$$\tilde{p} = (1 - p) \cdot \mathbf{I}\{\hat{p}_m < 0.5\} + p \cdot \mathbf{I}\{\hat{p}_m \geqslant 0.5\}$$

$$\mathcal{L}\{S_k\} \to \mathcal{N}(\tilde{p}, \tilde{p}(1 - \tilde{p})/k), \quad k \to \infty.$$

Мощность критерия на базе МП-предиктора

Теорема 2. Пусть верна $\mathcal{H}_1$ (7) и используется МП-предиктор. Тогда в асимптотике $m, k \to \infty$ критерий (6) является состоятельным, его мощность имеет вид:

$$\left| W_{m,k} - \left(\left(\Phi(c\sqrt{m}) \right) \left(1 - \Phi(\frac{\Delta}{2\sqrt{p(1-p)}} - c\sqrt{k}) \right) + \right. \right.$$
$$\left. \left. + \left(1 - \Phi(c\sqrt{m}) \right) \left(1 - \Phi(\frac{\Delta}{2\sqrt{p(1-p)}} + c\sqrt{k}) \right) \right) \right| \to 0,$$

where $c = (0.5 - p)/\sqrt{p(1-p)}$.

Построение предиктора Лемпеля-Зива

Предиктор Лемпеля-Зива: $\mathcal{M}$ – стационарные эргодические марковские источники конечного порядка

LZ-предиктор строится по последовательности, разбивая ее на непересекающиеся слова, каждое из которых не встречалось ранее

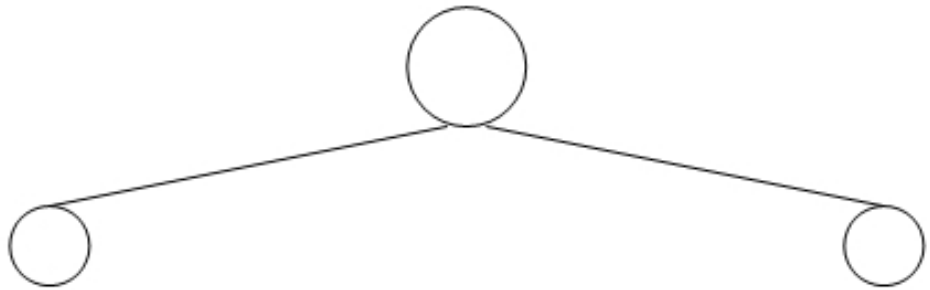
Проблема в NIST 800-22: использование оценок математического ожидания и дисперсии словаря вместо теоретических значений

Словарь LZ-предиктора представляется в виде дерева

Покажем на примере: $X = 1100101000100100$

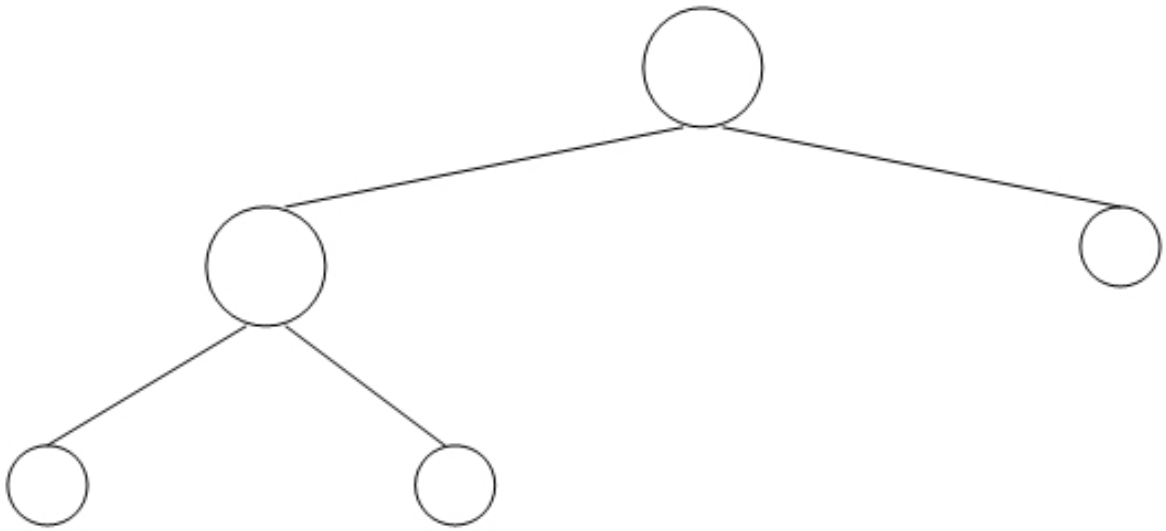
Построение дерева Лемпеля-Зива

Пример: $X_{-16}^{0} = 1100101000010000100$ Первоначальное дерево



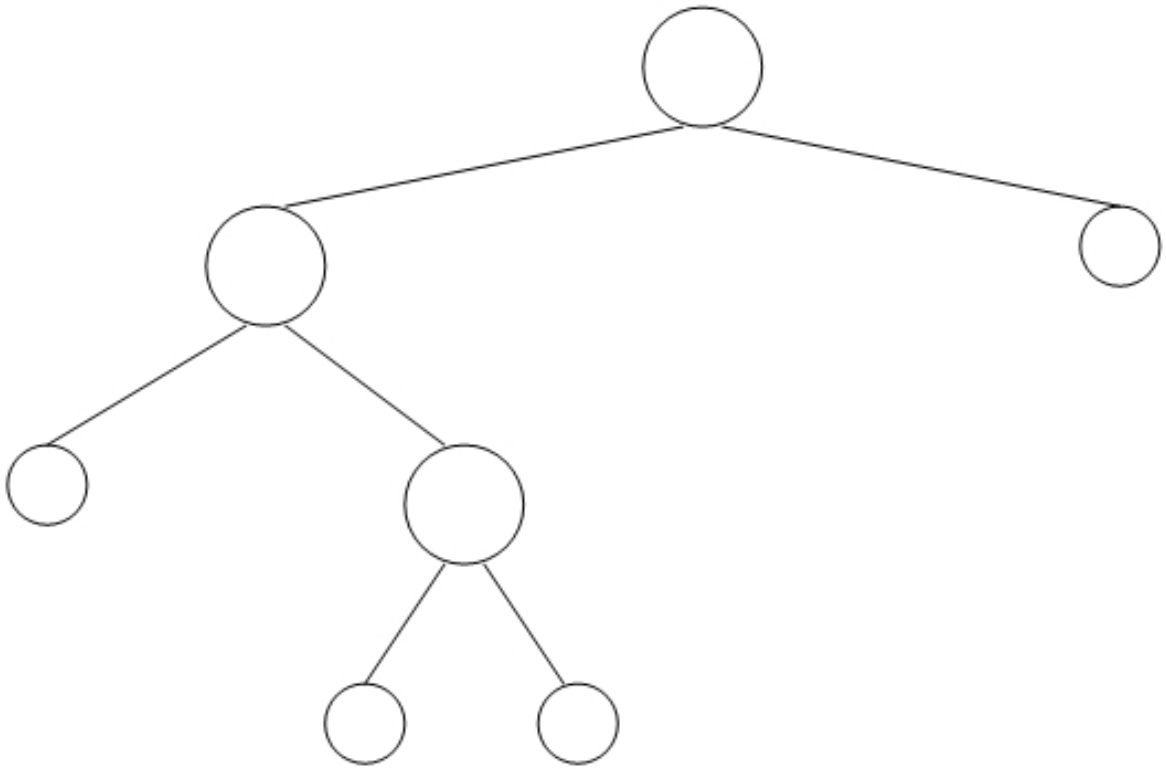
Построение дерева Лемпеля-Зива

Пример: $X_{-16}^{0} = 1100101000100100$



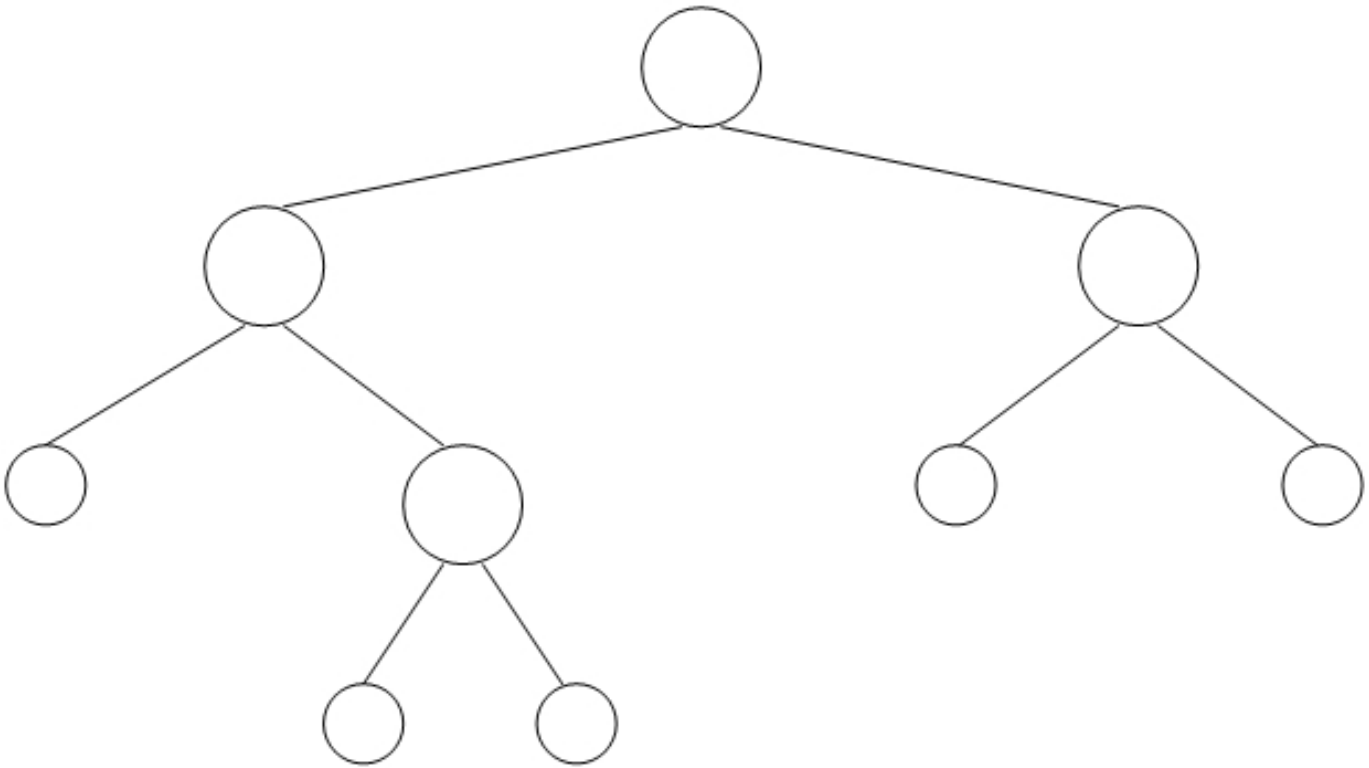
Построение дерева Лемпеля-Зива

Пример: $X_{-16}^{0} = 1100101000010001000100$



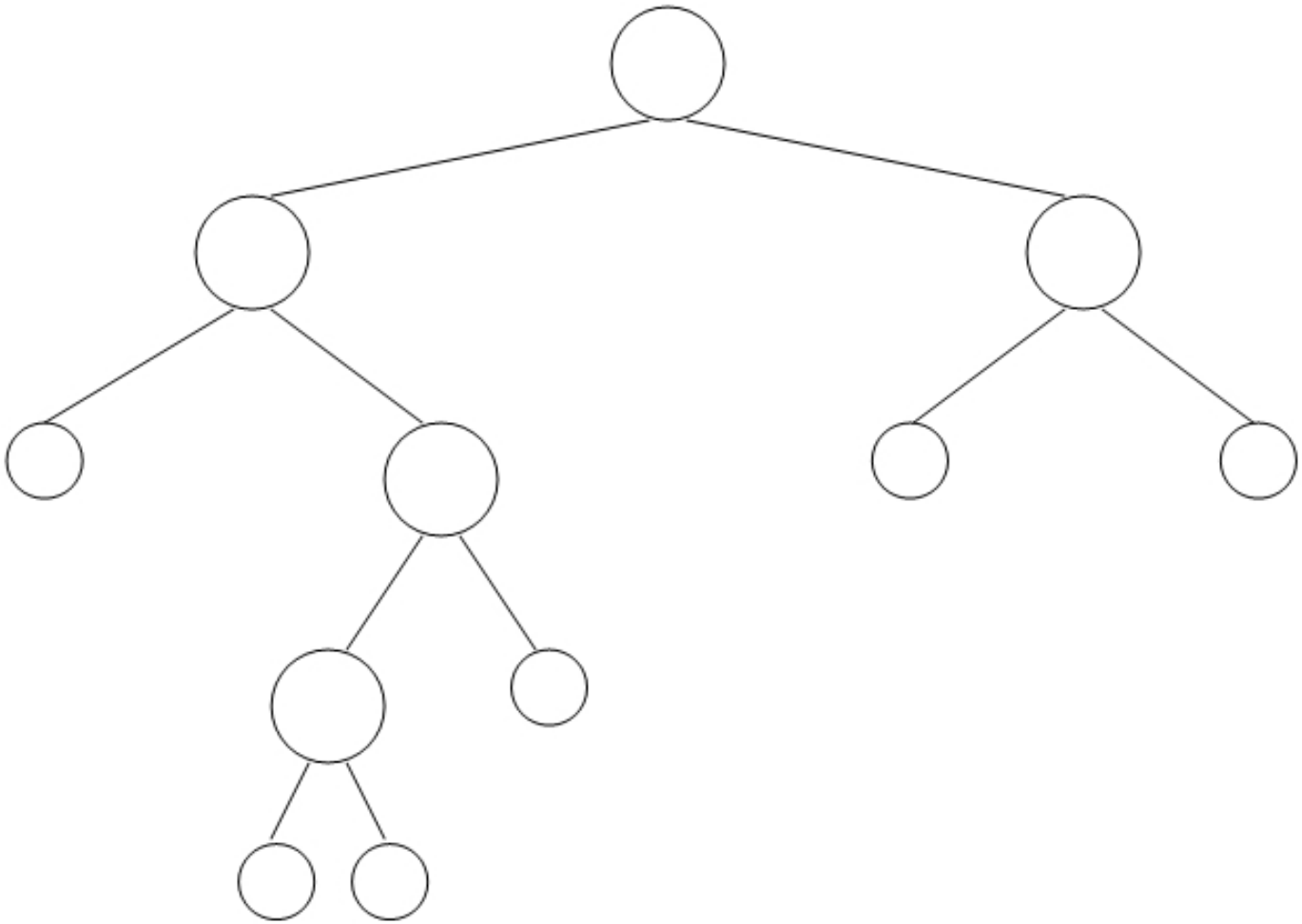
Построение дерева Лемпеля-Зива

Пример: $X_{-16}^{0} = 1100101000010001000100$



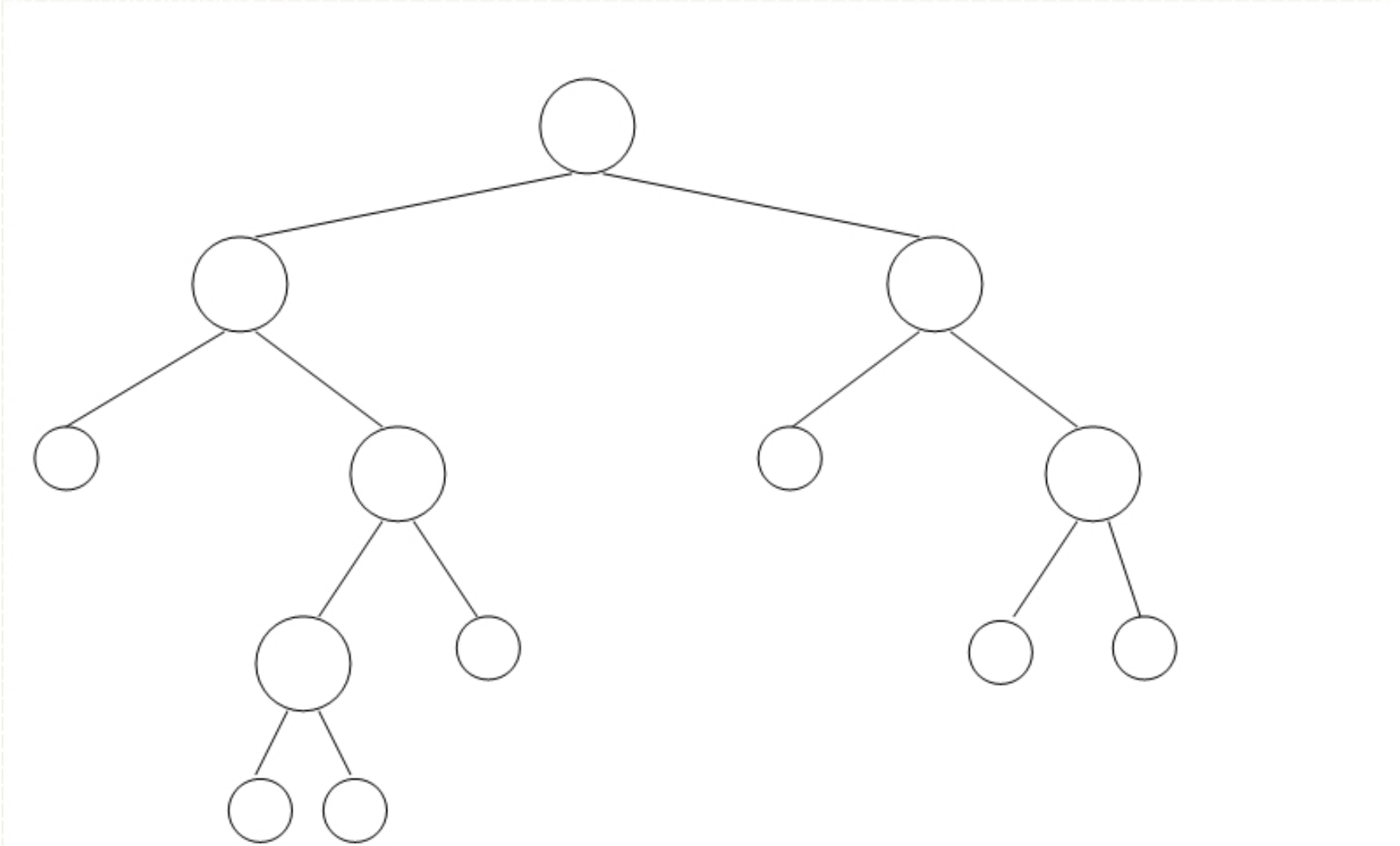
Построение дерева Лемпеля-Зива

Пример: $X^0_{-16} = 1100101010001000100$



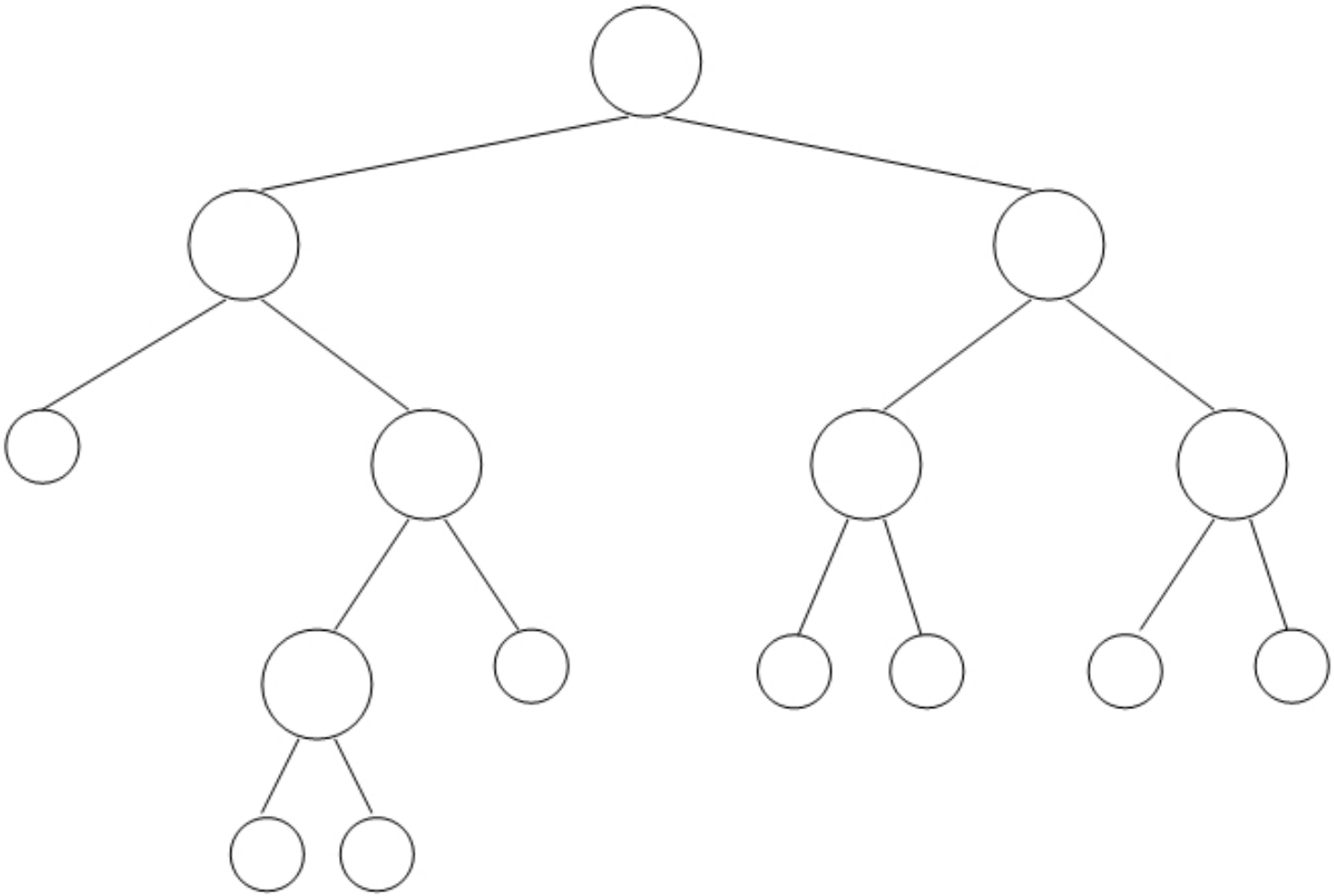
Построение дерева Лемпеля-Зива

Пример: $X_{-16}^{0} = 1100101000001000100$



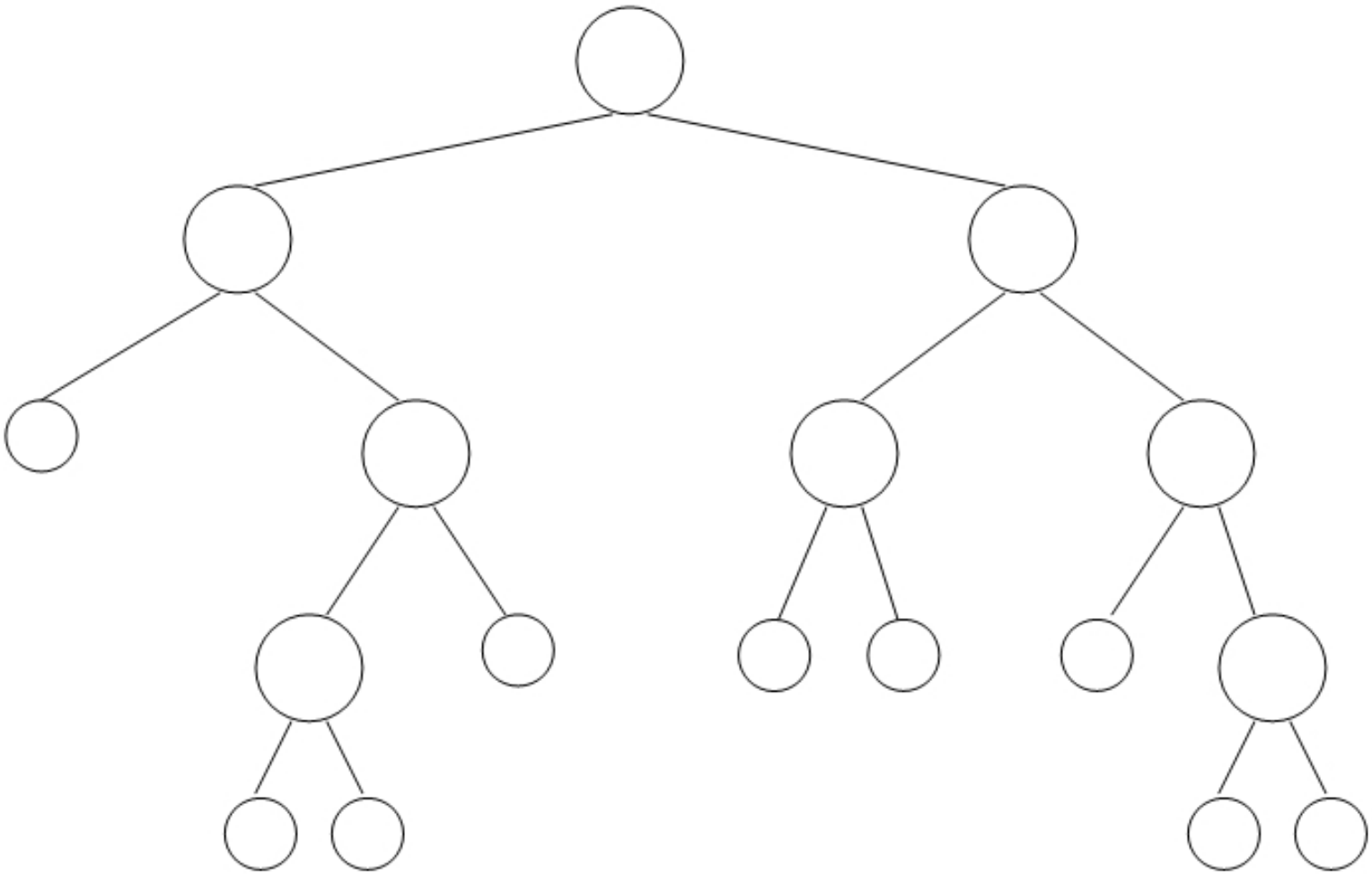
Построение дерева Лемпеля-Зива

Пример: $X_{-16}^{0} = 1100101000101000100$



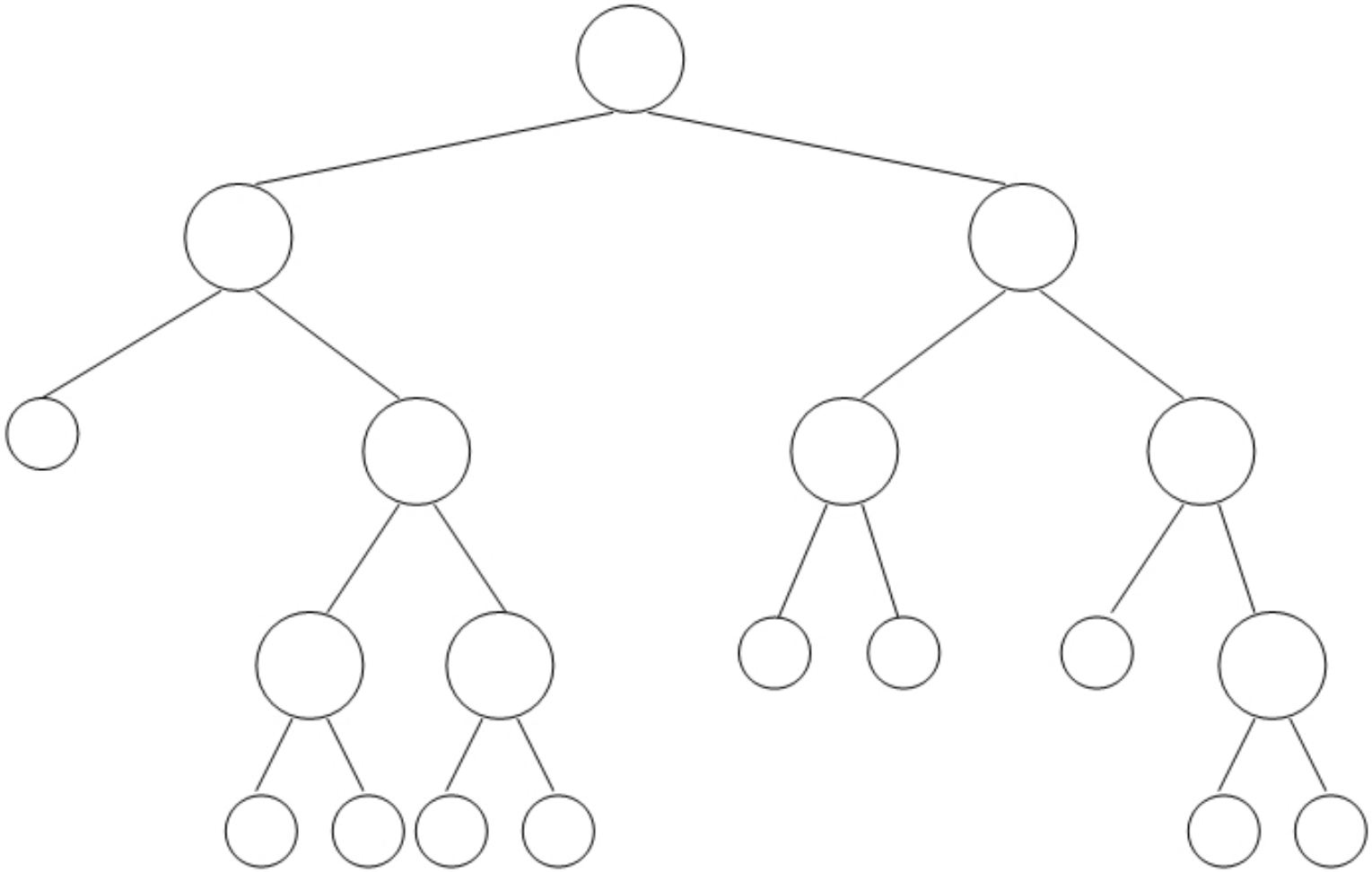
Построение дерева Лемпеля-Зива

Пример: $X^0_{-16} = 1100101000010000100$



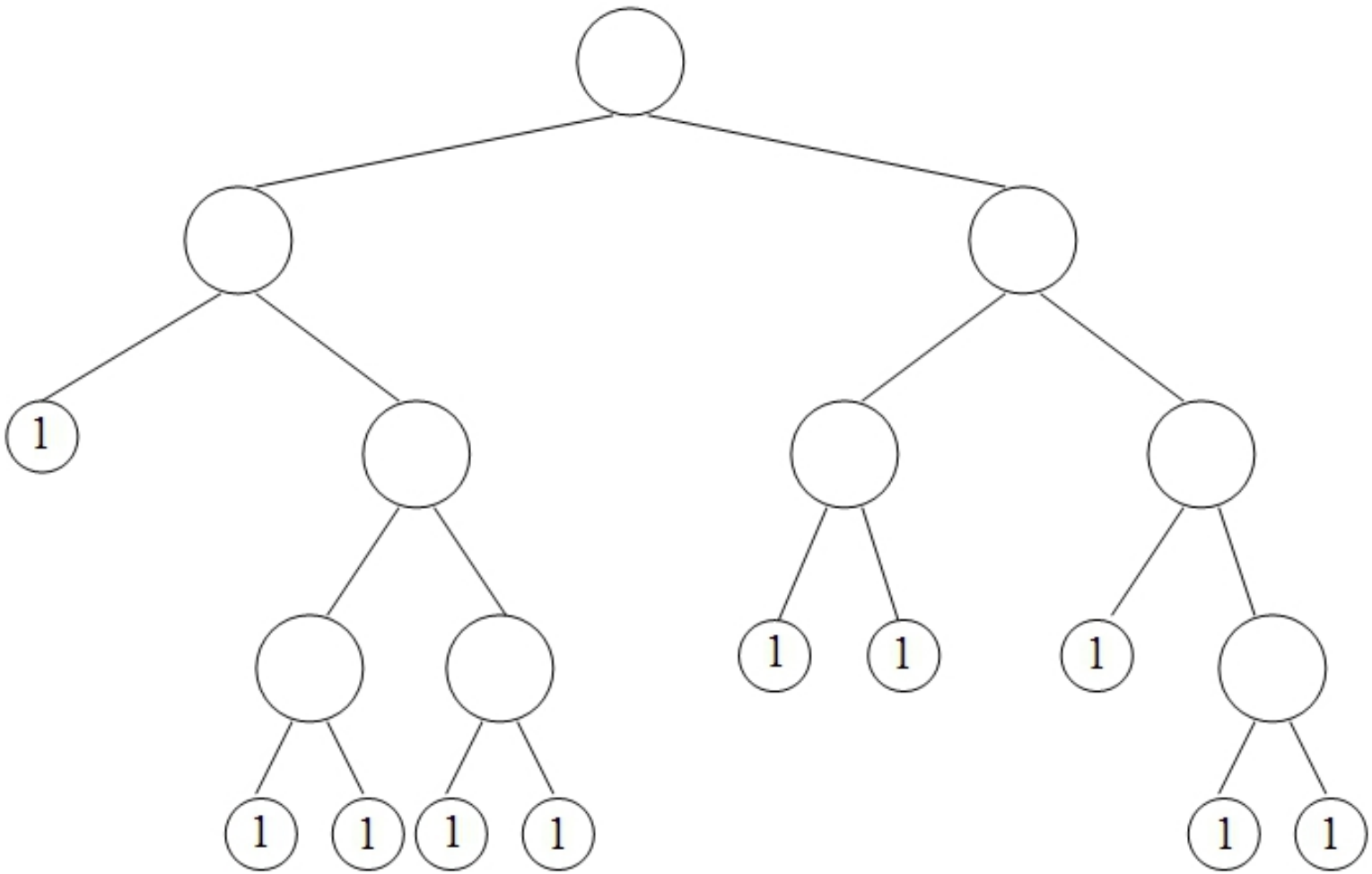
Построение дерева Лемпеля-Зива

Пример: $X^0_{-16} = 1100101000100010100$



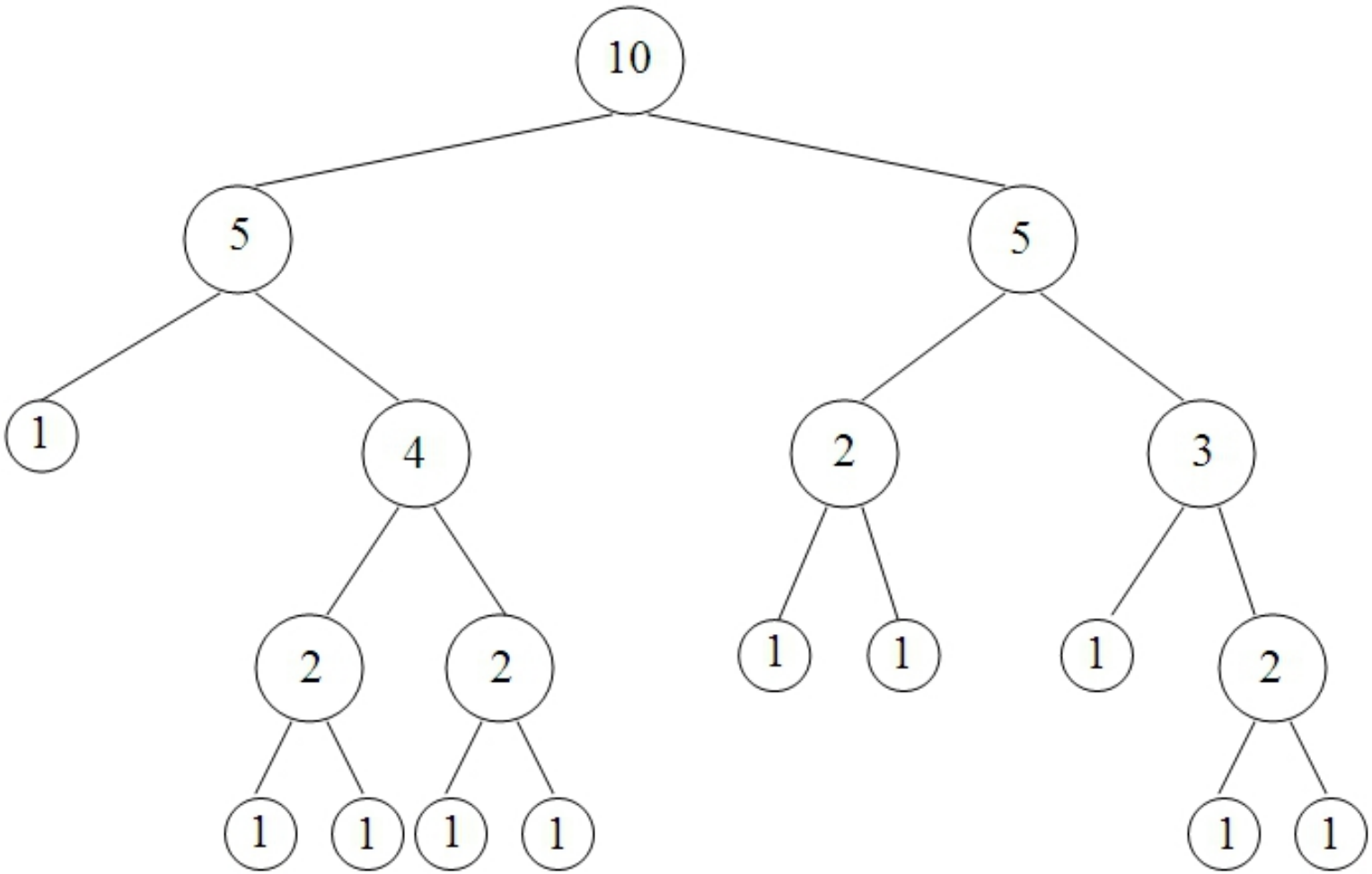
Построение дерева Лемпеля-Зива

Пример: $X^0_{-16} = 1100101000010001000100$. Расстановка весов



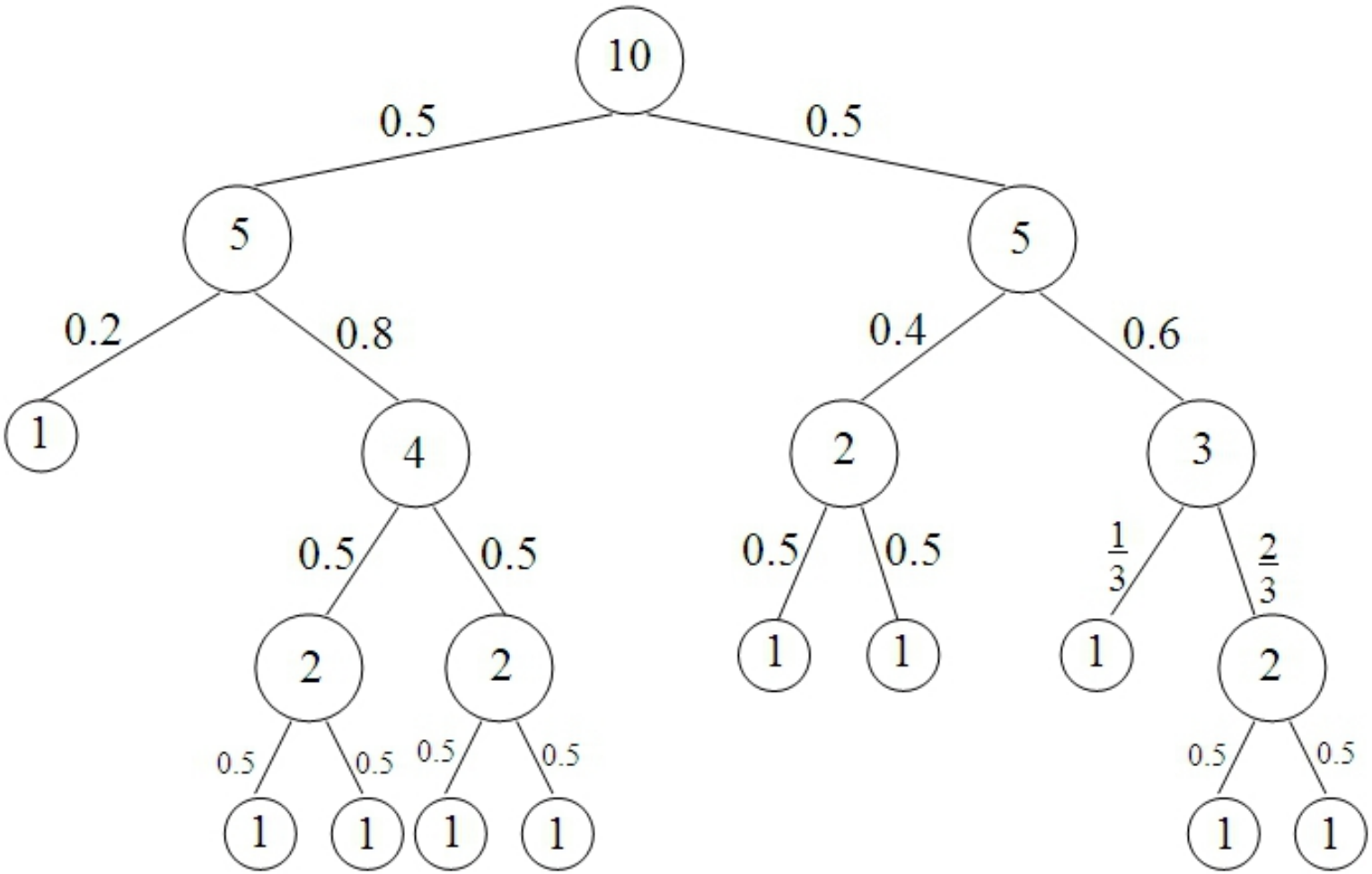
Построение дерева Лемпеля-Зива

Пример: $X^0_{-16} = 1100101000010000100$. Расстановка весов



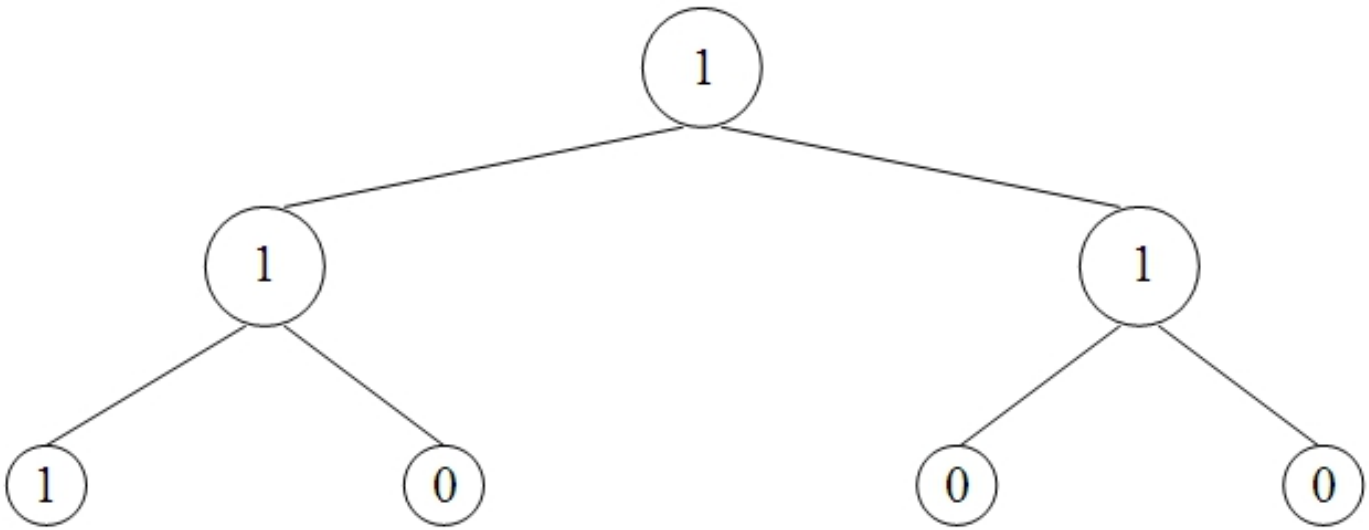
Построение дерева Лемпеля-Зива

Пример: $X_{-16}^{0} = 1100101000010000100$. Оценка вероятностей перехода.



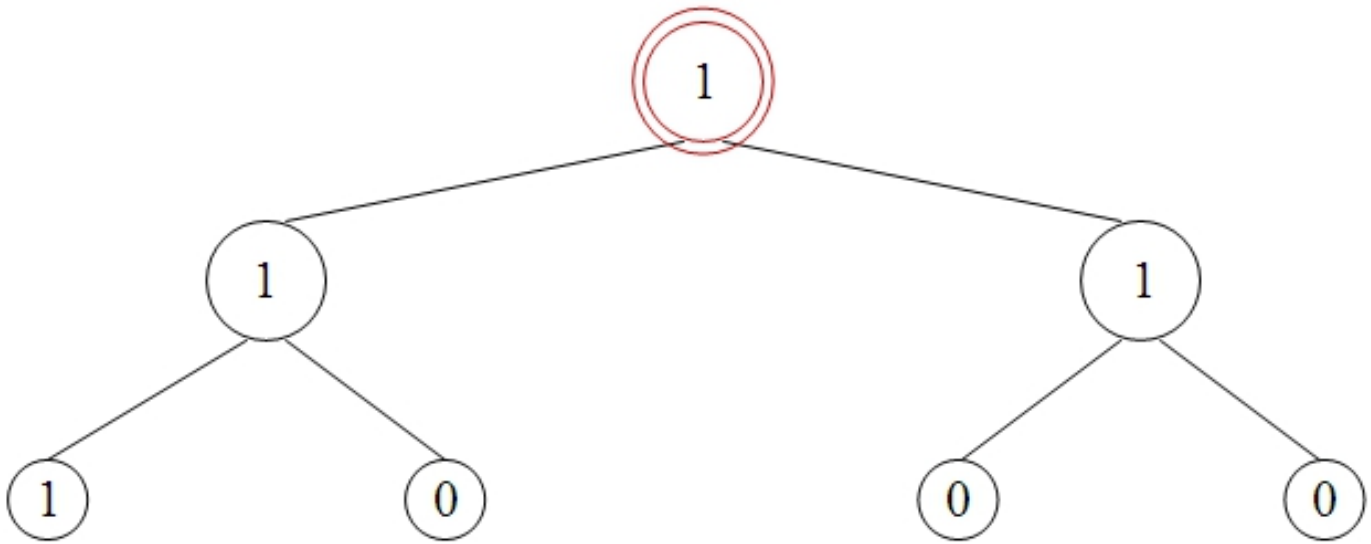
Использование предиктора для прогнозирования

Пример: $X_1^6 = 110010$



Использование предиктора для прогнозирования

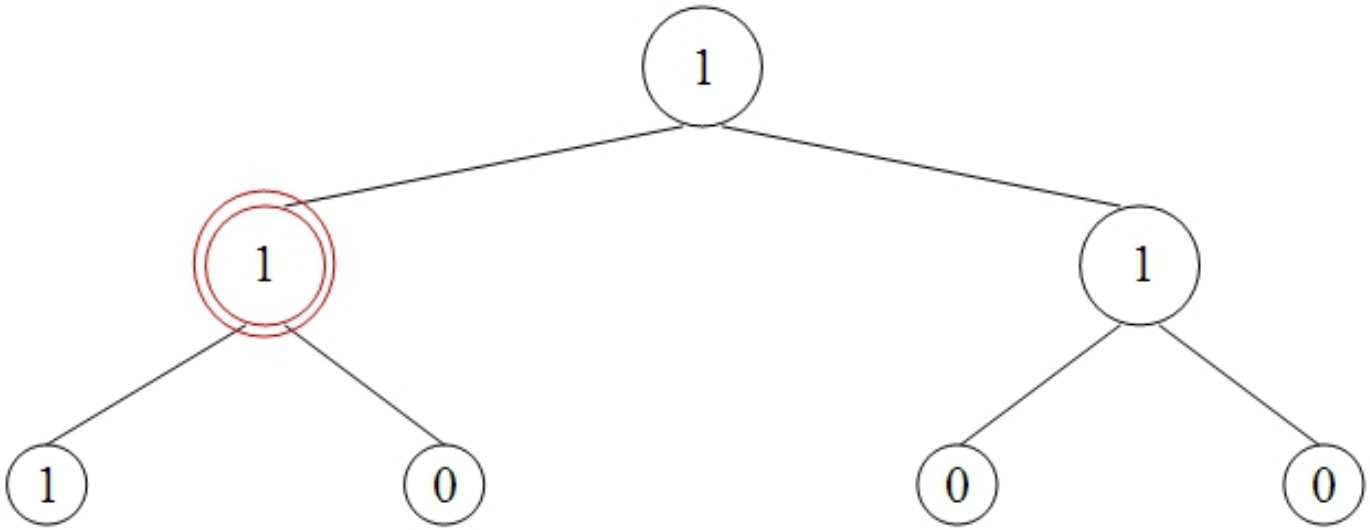
Пример: $X_1^6 = 110010$



$X_1 =$ “1” $\quad \hat{X}_1(s =$ “λ”$) =$ “1” $\quad Y_1 =$ “1”

Использование предиктора для прогнозирования

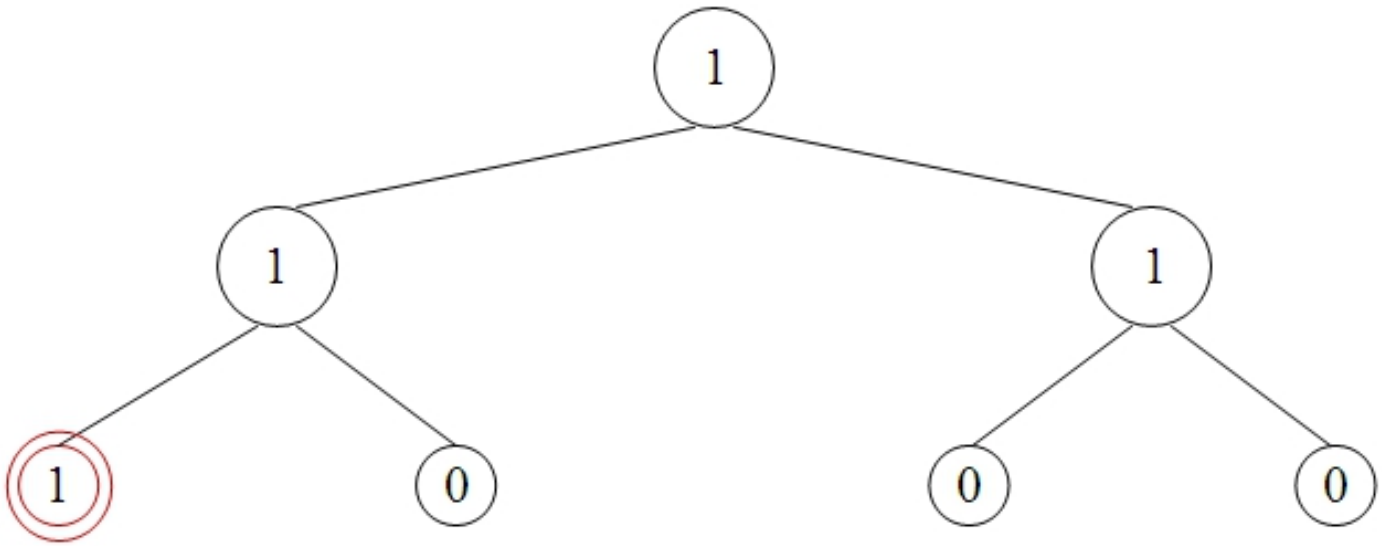
Пример: $X_1^6 = 110010$



$X_2 =$ “1” $\quad \hat{X}_2(s =$ “1”$) =$ “1” $\quad Y_2 =$ “1”

Использование предиктора для прогнозирования

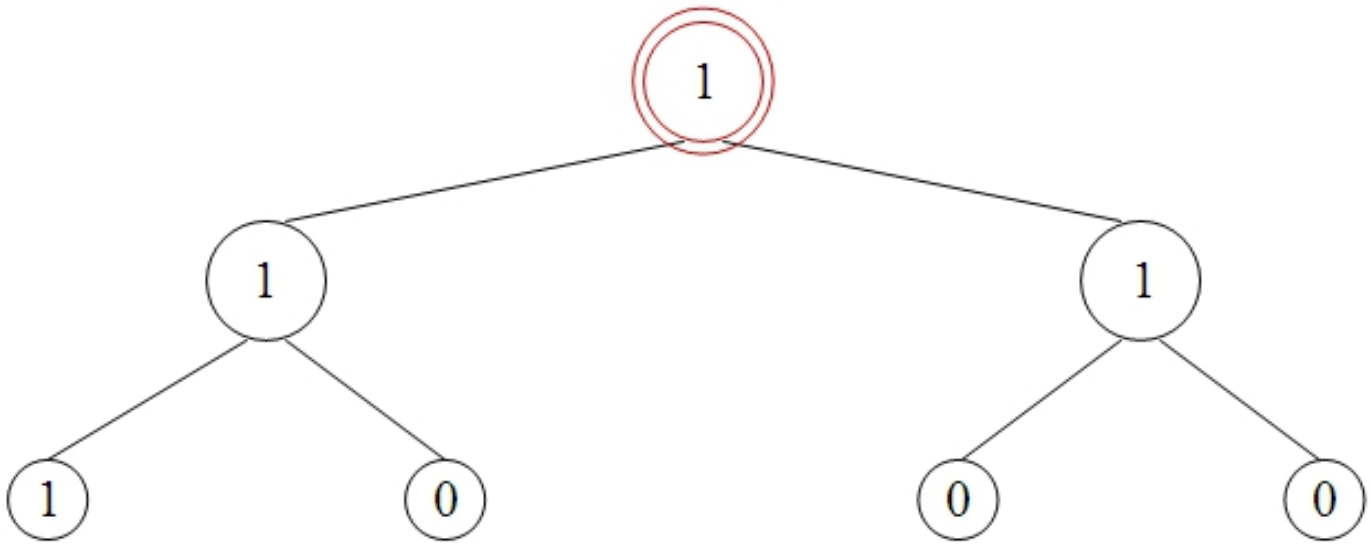
Пример: $X_1^6 = 110010$



$X_3 =$ “0” $\quad \hat{X}_3(s =$ “11”$) =$ “1” $\quad Y_3 =$ “0”

Использование предиктора для прогнозирования

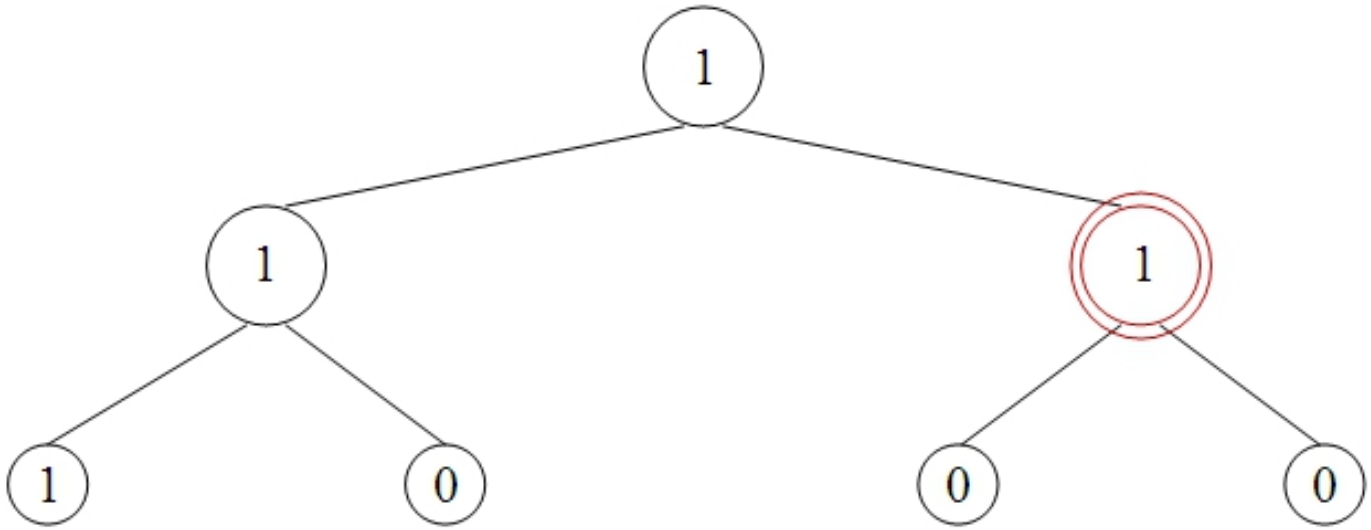
Пример: $X_1^6 = 110010$



$X_4 =$ “0” $\quad \hat{X}_4(s =$ “λ”$) =$ “1” $\quad Y_4 =$ “0”

Использование предиктора для прогнозирования

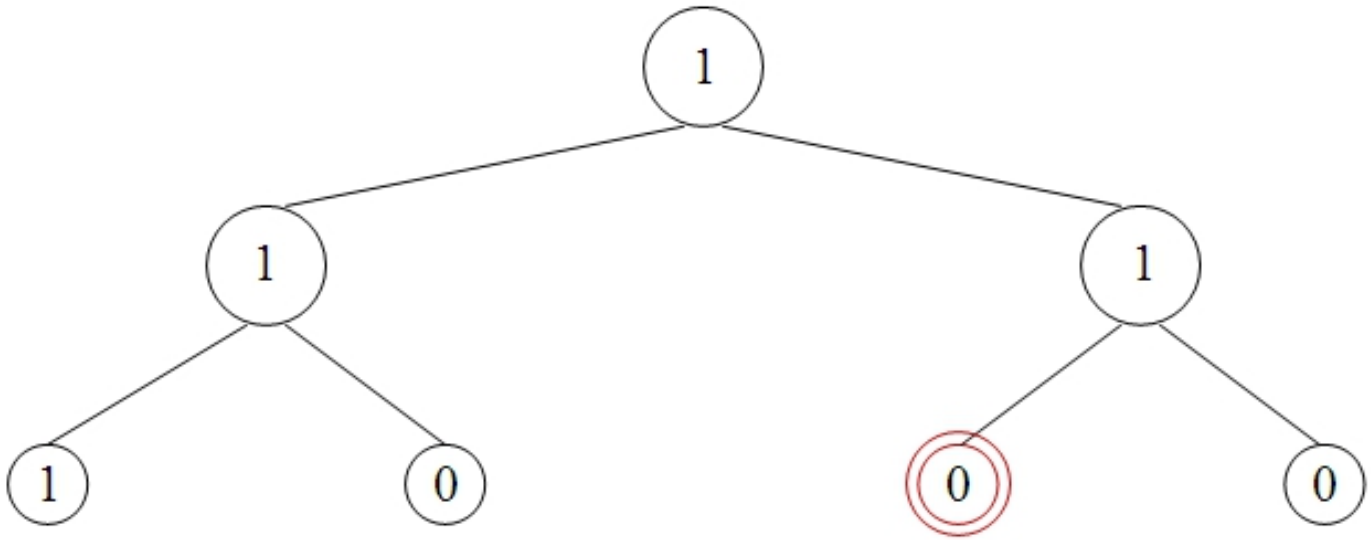
Пример: $X_1^6 = 110010$



$X_5 =$ “1” $\quad \hat{X}_5(s =$ “0”$) =$ “1” $\quad Y_5 =$ “1”

Использование предиктора для прогнозирования

Пример: $X_1^6 = 110010$



$X_6 =$ “0” $\quad \hat{X}_6(s =$ “01”$) =$ “0” $\quad Y_6 =$ “1”

Мощность критерия на базе предиктора Лемпеля-Зива

Модель Y_t – последовательность с.в. Бернулли, распределение вероятности успеха $\tilde{p}(s)$ определяется контекстом s

$$\tilde{p}(s) = ((1-p)\cdot \mathbf{I}\{\hat{X}^{(s)} = 0\} + p \cdot \mathbf{I}\{\hat{X}^{(s)} = 1\}).$$

Индикаторы $\{Y_t\}$ можно разделить на группы в зависимости от длины контекста, используемого при прогнозировании:

$$S_k = \frac{1}{k}\sum_{t=1}^{k} Y_t = \frac{1}{k}\sum_{i=0}^{L} S^{(i)}, \quad S^{(i)} = \sum_{t=0}^{k/(L+1)-1} Y_{1+i+(L+1)t}.$$

Теорема 2. Пусть верна гипотеза (7). Тогда для заданного дерева Лемпеля-Зива статистика $S^{(i)}$ имеет следующие моменты:

$$\mathbf{E}\left\{S^{(i)}\right\} = k\mu_{(i)}/(L+1), \quad \mathbf{D}\left\{S^{(i)}\right\} = k\mu_{(i)}(1-\mu_{(i)})/(L+1),$$

$$\mu_{(i)} = \sum_{s\in C: l(s)=i} \mathbf{P}\{s\}\,\tilde{p}(s), \quad \mathbf{P}\{s\} = p^{\sum_j s_j}(1-p)^{(i-\sum_j s_j)}.$$

Мощность критерия на базе предиктора Лемпеля-Зива

Теорема 3. Пусть верна гипотеза (7). Тогда в асимптотике $k \to \infty$ мощность критерия (6) на базе предиктора Лемпеля-Зива, построенного по некоторой последовательности X^0_{-m+1}, имеет следующее асимптотическое выражение:

$$\left| W_k(X^0_{-m+1}) - \left(1 - \Phi\left(\frac{\Delta}{2\sqrt{n}\sqrt{\sigma}} + \frac{0.5-\mu}{\sqrt{\sigma}}\right)\right)\right|,$$

$$\mu = \frac{1}{(L+1)}\sum_{i=0}^{L}\mu_{(i)}, \qquad \sigma = \frac{1}{k(L+1)}\left(\sum_{i=0}^{L}\mu_{(i)}(1-\mu_{(i)}) + \right.$$

$$\left. +2\sum_{j>i}\sum_{s\in C: l(s)=j}\mathbf{P}\{s\}\,\tilde{p}(s)\left(\mathbf{I}\{\hat{X}(s_1^i) = s_{i+1}\} - \mu_{(i)}\right)\right).$$

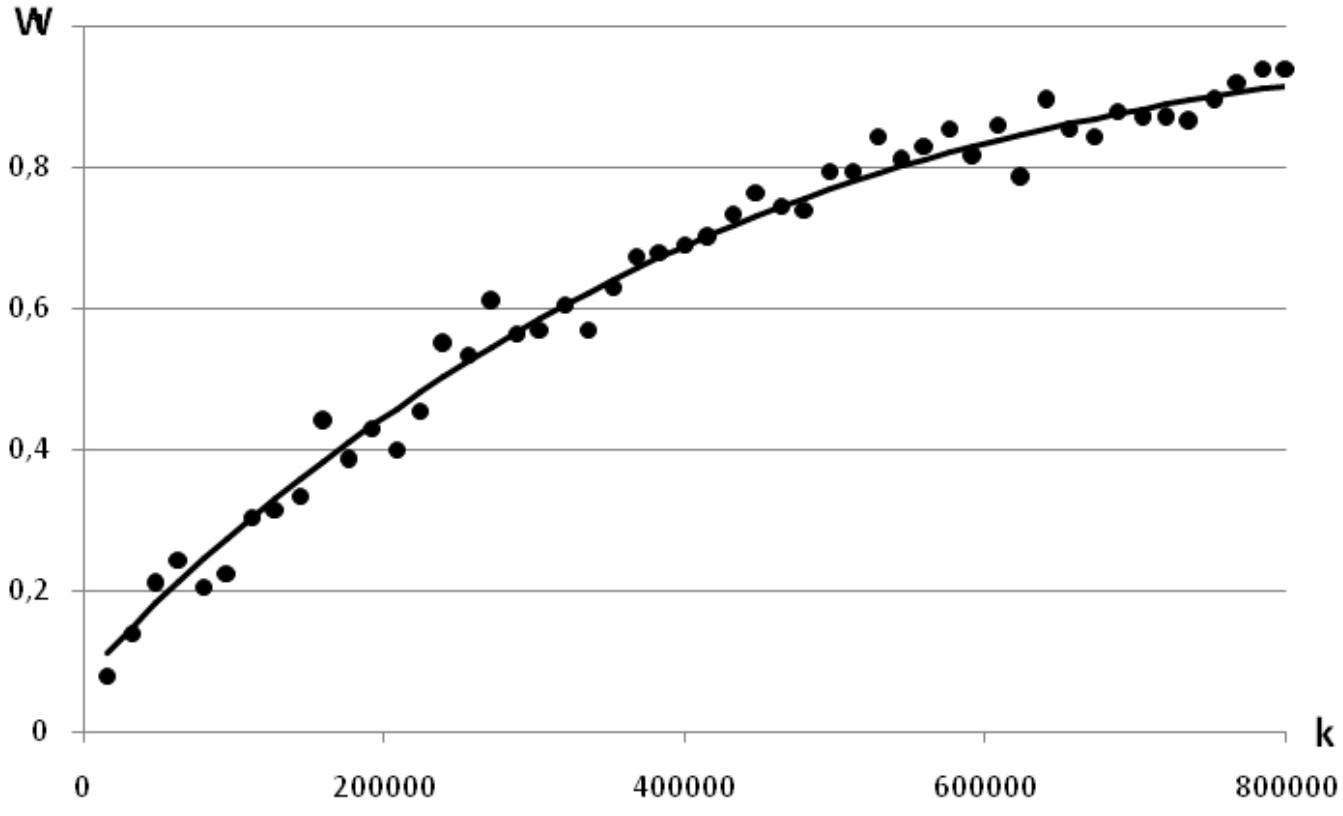
Вычислительный эксперимент

Альтернатива $\mathcal{H}_1$ – испытания Бернулли : $p = 0.503, \quad n = 10^6$

$m = 2 \cdot 10^5$, $k = 8 \cdot 10^5$

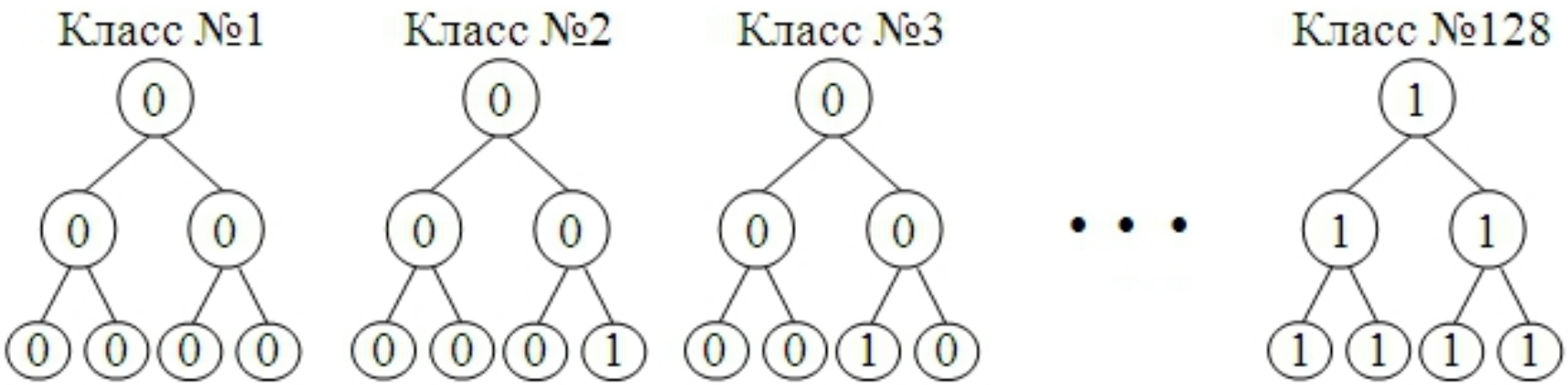
теоретическая мощность(——)

оценка мощности методом Монте-Карло(•)



Анализ числа классов

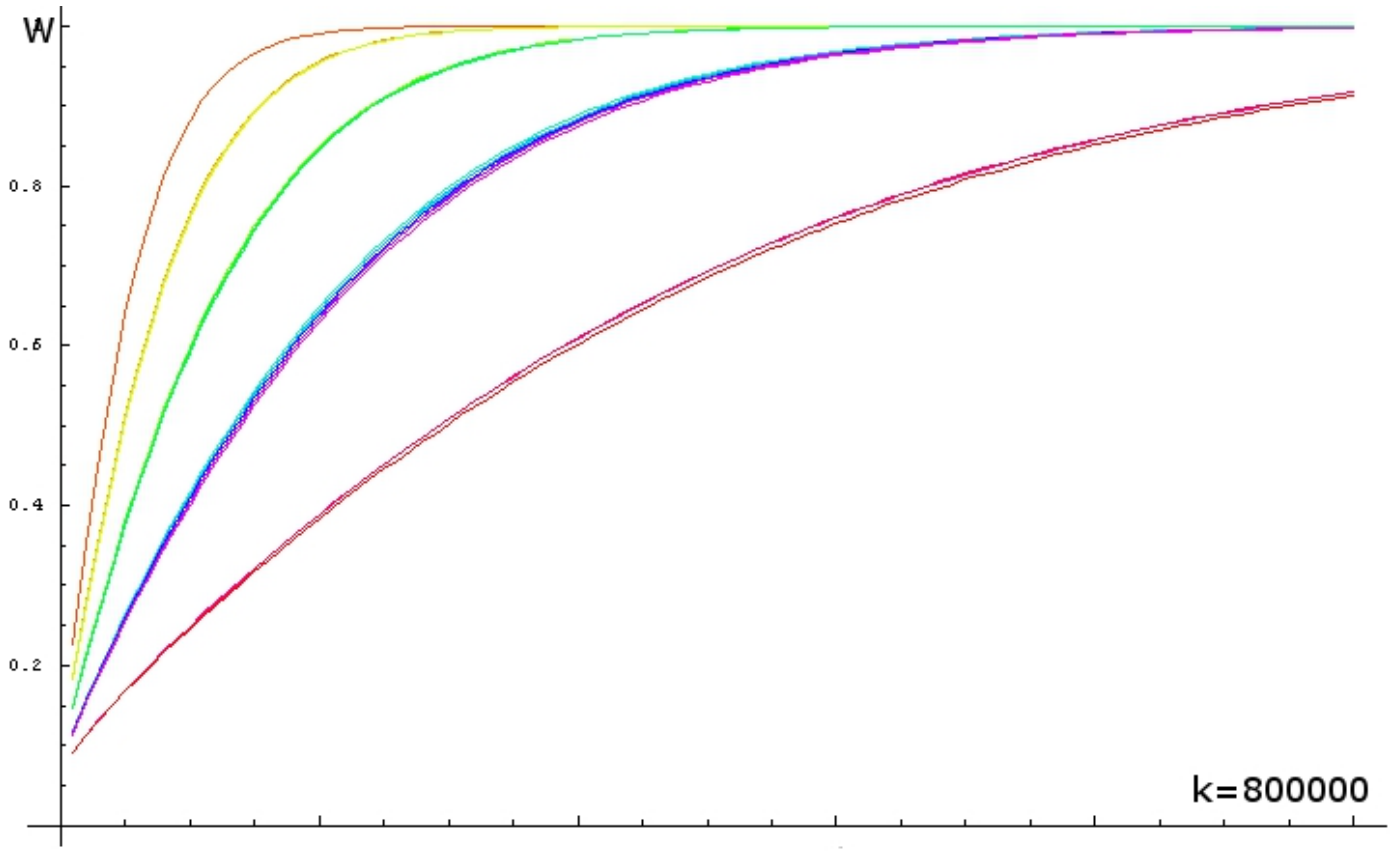
Рассмотрим $\mathcal{H}_1$ (7) с $p > 0.5$
Оптимальный МП-предиктор $\hat{X}_t = 1$
LZ-предиктор, построенный по X^0_{-m+1}, принадлежит одному из классов эквивалентности
Класс определяется множеством $\{\hat{X}^{(s)}\}, s \in LZ - tree$. Например, для $L = 2$



С ростом L число классов растет экспоненциально!

Анализ числа классов

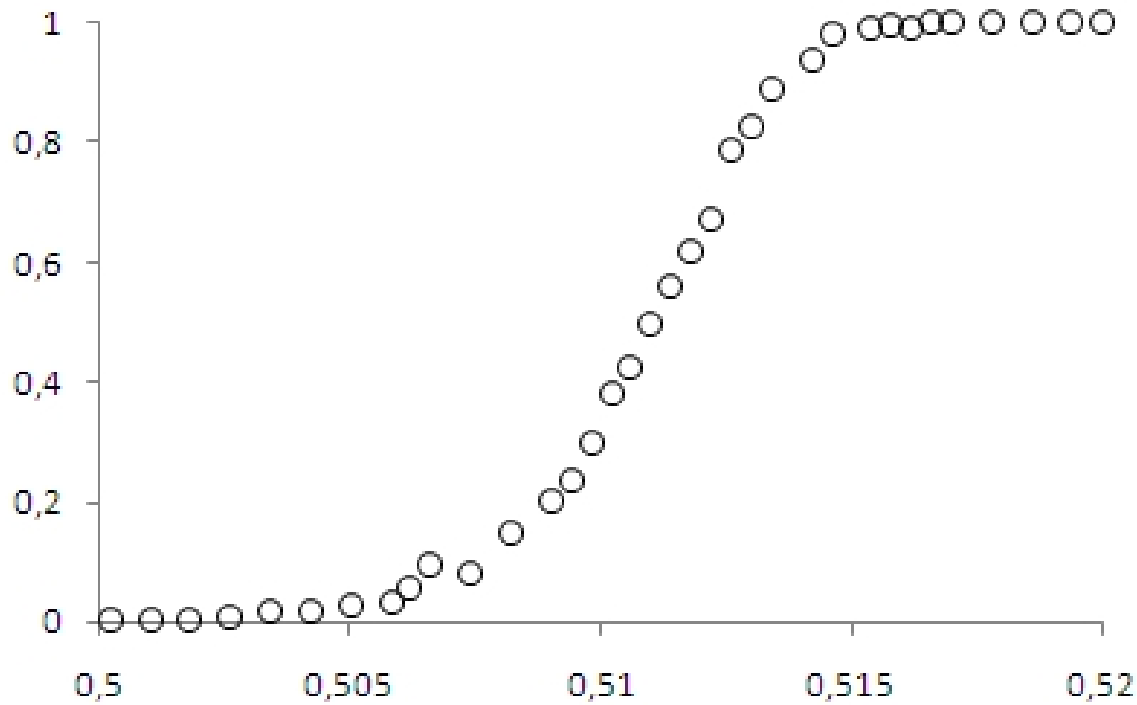
Могут быть получены границы для мощности критерия на базе предиктора Лемпеля-Зива



Сравнение мощностей критерия на базе предиктора Лемпеля-Зива и критерия из NIST SP 800-22

Альтернатива $\mathcal{H}_1 -$ испытания Бернулли : $p = \frac{1}{2} + \varepsilon, \quad n = 10^6.$

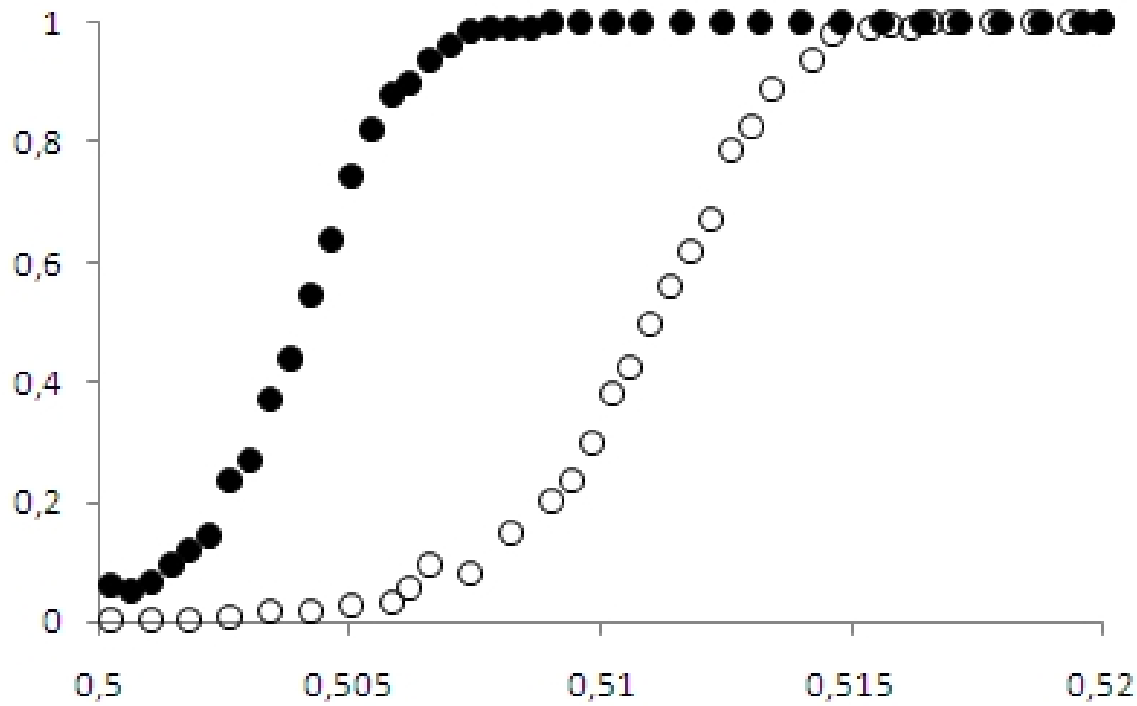
мощность критерия сжатия Лемпеля-Зива из NIST SP 800-22($\circ$)
мощность критерия на базе предиктора Лемпеля-Зива($\bullet$)
мощность критерия на базе МП-предиктора (——)



Сравнение мощностей критерия на базе предиктора Лемпеля-Зива и критерия из NIST SP 800-22

Альтернатива $\mathcal{H}_1 -$ испытания Бернулли : $p = \frac{1}{2} + \varepsilon, \quad n = 10^6.$

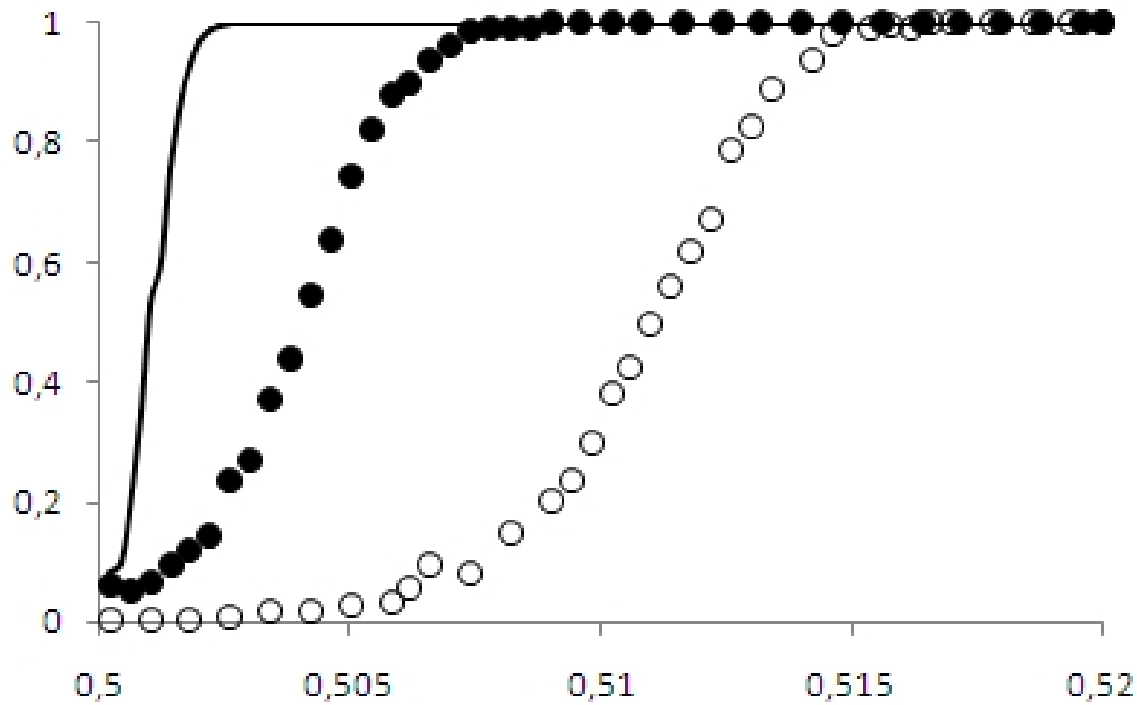
мощность критерия сжатия Лемпеля-Зива из NIST SP 800-22($\circ$)
мощность критерия на базе предиктора Лемпеля-Зива($\bullet$)
мощность критерия на базе МП-предиктора (——)



Сравнение мощностей критерия на базе предиктора Лемпеля-Зива и критерия из NIST SP 800-22

Альтернатива $\mathcal{H}_1 -$ испытания Бернулли : $p = \frac{1}{2} + \varepsilon, \quad n = 10^6.$

мощность критерия сжатия Лемпеля-Зива из NIST SP 800-22($\circ$)
мощность критерия на базе предиктора Лемпеля-Зива($\bullet$)
мощность критерия на базе МП-предиктора (——)



Заключение

- Приведен подход к построению критерия тестирования гипотезы о случайности бинарной последовательности на базе любого универсального предиктора
- Предложена двухэтапная процедура и исследованы ее свойства в случае модели испытаний Бернулли и критерия на базе предиктора максимального правдоподобия
- Найдена мощность критерия на базе универсального предиктора Лемпеля-Зива
- Проведенные вычислительные эксперименты демонстрируют перспективность предложенного подхода
- Возможно использование универсальных предикторов для оценки энтропии

Заключение

- Приведен подход к построению критерия тестирования гипотезы о случайности бинарной последовательности на базе любого универсального предиктора
- Предложена двухэтапная процедура и исследованы ее свойства в случае модели испытаний Бернулли и критерия на базе предиктора максимального правдоподобия
- Найдена мощность критерия на базе универсального предиктора Лемпеля-Зива
- Проведенные вычислительные эксперименты демонстрируют перспективность предложенного подхода
- Возможно использование универсальных предикторов для оценки энтропии

Заключение

- Приведен подход к построению критерия тестирования гипотезы о случайности бинарной последовательности на базе любого универсального предиктора
- Предложена двухэтапная процедура и исследованы ее свойства в случае модели испытаний Бернулли и критерия на базе предиктора максимального правдоподобия
- Найдена мощность критерия на базе универсального предиктора Лемпеля-Зива
- Проведенные вычислительные эксперименты демонстрируют перспективность предложенного подхода
- Возможно использование универсальных предикторов для оценки энтропии

Заключение

- Приведен подход к построению критерия тестирования гипотезы о случайности бинарной последовательности на базе любого универсального предиктора
- Предложена двухэтапная процедура и исследованы ее свойства в случае модели испытаний Бернулли и критерия на базе предиктора максимального правдоподобия
- Найдена мощность критерия на базе универсального предиктора Лемпеля-Зива
- Проведенные вычислительные эксперименты демонстрируют перспективность предложенного подхода
- Возможно использование универсальных предикторов для оценки энтропии

Заключение

- Приведен подход к построению критерия тестирования гипотезы о случайности бинарной последовательности на базе любого универсального предиктора
- Предложена двухэтапная процедура и исследованы ее свойства в случае модели испытаний Бернулли и критерия на базе предиктора максимального правдоподобия
- Найдена мощность критерия на базе универсального предиктора Лемпеля-Зива
- Проведенные вычислительные эксперименты демонстрируют перспективность предложенного подхода
- Возможно использование универсальных предикторов для оценки энтропии