распределения вероятности $\mathscr{R}$ до распределения вероятности $\mathscr{P} * Q$.

Затем лектор ввел еще одно новое для нас понятие: код с переменной длиной слова. Пусть ξ — случайная переменная, принимающая значения x_1, x_2 ... ..., x_N с вероятностью $p_1, p_2, \ldots, p_N$. Сначала мы рассмотрели наиболее простой случай — так называемый *двоичный код*, в котором ξ принимает лишь два значения: 0 и 1. Закодируем значения $x_1, x_2, \ldots, x_N$ различными наборами (длина которых может варьироваться) нулей и единиц. Если ни одно кодовое слово не совпадает с начальным отрезком какого-нибудь другого кодового слова, то *код* называется *префиксным.* Префиксные коды обладают тем большим преимуществом, что позволяют не указывать конец слова и записывать одно кодовое слово за другим без пробела, поскольку префиксность кода позволяет разбить «сплошной» текст на отдельные слова одним и только одним способом. Иначе говоря, префиксный код допускает однозначное декодирование. Разумеется, далеко не всякий код, допускающий однозначное декодирование, префиксный (чтобы убедиться в этом, достаточно рассмотреть код, состоящий из двух слов: 0 и 01). В подтверждение своих слов о преимуществе префиксных кодов лектор привел пример. Пусть ξ — случайная величина, принимающая значения 0, 1, 2, ..., 9. Закодируем цифры от 0 до 9 наборами нулей и единиц следующим образом:

$$
\begin{aligned}
0 &\to 00,\\
1 &\to 01,\\
2 &\to 1000,\\
3 &\to 1001,\\
4 &\to 101,\\
5 &\to 110,\\
6 &\to 1110,\\
7 &\to 11110,\\
8 &\to 111110,\\
9 &\to 111111
\end{aligned}
$$

Нетрудно видеть, что перед нами префиксный код. Если кодовые слова записать без пробелов одно за другим в произвольной последовательности (кодовое слово может встречаться в тексте любое число раз), то полученный набор знаков можно разбить на кодовые слова одним и только одним способом. Следовательно, кодированная запись допускает однозначное декодирование. Например, набор нулей и единиц

$$1110101111101000000001100111111110101$$

допускает лишь одно разбиение на кодовые слова

$$1110|101|11110|1000|00|00|01|1001|111111|01|01$$

Декодировать его можно только как

$$64720013911.$$

Префиксный код называется примитивным, если его невозможно сократить, то есть если при вычеркивании любого знака хотя бы в одном кодовом слове код перестает быть префиксным. Нетрудно видеть, если код префиксный и мы выбрали любой набор из s нулей и единиц, не входящий в число кодовых слов, то возможно одно из двух: либо не существует кодового слова, начальный отрезок которого совпадает с нашим набором, либо (если такое кодовое слово существует), приписав к концу нашего набора нуль или единицу, мы получим какое-то кодовое слово или начальный отрезок кодового слова. Примитивный (двоичный) префиксный код можно изобразить в виде особого графа — дерева, у которого из корня и любой вершины, кроме свободных, исходят по две ветви. Если мы условимся всегда сопоставлять левой ветви нуль, а правой единицу, то каждой свободной вершине дерева будет однозначно соответствовать некий набор нулей и единиц, показывающий, в какой последовательности нужно сворачивать направо и налево, добираясь до этой вершины из корня дерева. Итак, предложенный нами способ кодирования позволяет каждой свободной вершине дерева поставить в соответствие определенное кодовое слово. Построенный граф называется *деревом кода*. Например, примитивному префиксному коду, рассмотренному на этой лекции, соответствует дерево, изображенное на рис. 3.

Пусть N_k — число кодовых k-значных слов в примитивном префиксном коде. Если r — число знаков в самом длинном кодовом слове, то справедливо тождество