InT 1. Докажите, что:

а) код Шеннона–Фано является префиксным;

б) если центральный отрезок относить туда, куда попала его большая часть, то кодирование Шеннона–Фано не является сбалансированным (то есть не существует константы d, для которой выполнено $\ell(c_i) < -\log p_i + d$ для любых k и любых исходных вероятностей $p_1, \dots, p_k$);

в) если центральный отрезок всегда относить к правой половине, то кодирование Шеннона–Фано также не является сбалансированным.

InT 2. Докажите, что арифметическое кодирование сбалансировано с константой 2.

InT 3. Приведите пример такого распределения вероятностей, что код Шеннона–Фано не является оптимальным.

InT 4. Пусть последовательность $\alpha \to \beta \to \gamma$ образует цепь Маркова. Докажите, что:

а) $I(\alpha{:}\,\gamma) \leq I(\alpha{:}\,\beta)$;

б) $I(\alpha{:}\,\gamma) \leq I(\beta{:}\,\gamma)$.

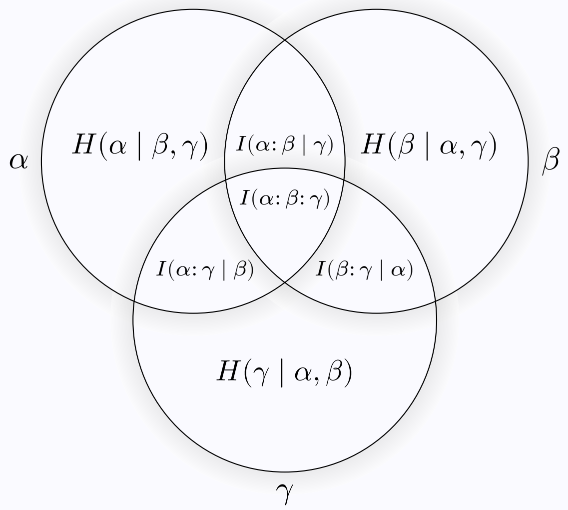
InT 5. Пусть $\{a_1, a_2, \dots, a_n\}$ — произвольный алфавит и $p_1, p_2, \dots, p_n$ — вероятности букв этого алфавита. Докажите, что для любого инъективного кодирования букв этого алфавита средняя длина кода не меньше $h - 2\log(h+1) - 2$, где h — энтропия распределения с вероятностями $p_1, p_2, \dots, p_n$.

Определение

Определим **общую информацию** трёх случайных величин:

$$I(\alpha{:}\,\beta{:}\,\gamma) := I(\alpha{:}\,\beta) - I(\alpha{:}\,\beta \mid \gamma).$$

Соотношения на информационные величины имеют удобную геометрическую интерпретацию. При помощи диаграмм Эйлера можно сопоставить площади каждой из получившихся замкнутых области некоторую информационную величину. В частности, прощать каждого круга соответствует энтропии указанной случайной величины.



InT 6. Постройте три таких случайных выличины α, β, γ, что $I(\alpha{:}\,\beta{:}\,\gamma) < 0$.

Определение

Коммуникационный протокол для функции $f: X \times Y \to Z$ — это корневое двоичное дерево, которое описывает совместное вычисление Алисой и Бобом функции f. В этом дереве каждая внутренняя вершина v помечена меткой a или b, означающей очередь хода Алисы или Боба соответственно. Для каждой вершины, помеченной a, определена функция $g_v: X \to \{0, 1\}$, которая говорит Алисе, какой бит нужно послать, если вычисление находится в этой вершине. Аналогично, для каждой вершины v с пометкой b определена функция $h_v: Y \to \{0, 1\}$, которая определяет бит, который Боб должен отослать в этой вершине. Каждая внутренняя вершина имеет двух потомков, ребро к первому потомку помечено 0, а ребро ко второму потомку помечено 1. Каждый лист помечен значением из множества Z.
Каждая пара входов (x, y) определяет путь от корня до листа в описанном двоичном дереве естественным обрзом. Будем говорить, что коммуникационный протокол вычисляет функцию f, если для всех пар $(x, y) \in X \times Y$, этот путь заканчивается в листе с пометкой $f(x, y)$.
Коммуникационной сложностью функции f назовем наименьшую глубину протокола, вычисляющего функцию f, и будем ее обозначать $\mathrm{D}(f)$.

InT 7. Пусть $f: X \times Y \to \{0, 1\}$. Рассмотрим матрицу M^f, строки которой соответствуют элементам множества X, столбцы элементам множества Y, а в ячейке $M_{x,y}$ написано значение $f(x, y)$. Покажите, что:

а) $\mathrm{rk}_{\mathbb{R}}(M_{x,y})$ не превосходит числа листьев в коммуникационном протоколе для f (*подсказка:* для вершины протокола опишите множество входов, которое приводит в нее);

б) $\mathrm{D}(\mathsf{EQ}_n) \geq n$, где $\mathsf{EQ}_n: \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ и $\mathsf{EQ}_n(x, y) = 1 \Leftrightarrow x = y$.

InT (S 1 P 3). Для множества $A \subseteq \mathbb{N}^4$ будем обозначать $\pi_{ijk}(A)$ проекцию A на координатную плокость, задаваемую осями i, j, k (индексы $i, j, k \in [4]$). Докажите, что для любого конечного A выполняется:

$$3 \log |A| \leq \log |\pi_{123}(A)| + \log |\pi_{124}(A)| + \log |\pi_{134}(A)| + \log |\pi_{234}(A)|.$$

InT (S 2 P 6). Имеется набор из n камней. Сколько взвешиваний необходимо, чтобы найти самый тяжелый и самый лёгкий камни (на каждую чашу можно класть не более одного камню)?