Энтропия Шеннона, часть вторая

Пусть задано совместное распределение вероятностей пары случайных величин (α, β). Пусть α принимает значения в конечном множестве $\{a_1, \ldots, a_n\}$, а β в конечном множестве $\{b_1, \ldots, b_m\}$ (распределение вероятностей есть набор чисел

$$p_{ij} = \mathrm{Prob}[\alpha = a_i,\ \beta = b_j]$$

для $i = 1 \ldots n$, $j = 1 \ldots m$).

Задача 1. Докажите, что $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$, причём равенство выполняется в том и только том случае, когда α и β независимы.

Для каждого фиксированного значения a_i можно рассмотреть условное распределение случайной величины β при заданном значении α:

$$p_{j|i} = \mathrm{Prob}[\beta = \beta_j \mid \alpha = a_i],$$

будем обозначать энтропию этого распределения $H(\beta \mid \alpha = a_i)$. Усреднение этих энтропий называется *условной энтропией* β относительно α:

$$H(\beta \mid \alpha) := \sum_i H(\beta \mid \alpha = a_i) \cdot \mathrm{Prob}[\alpha = a_i]$$

Задача 2. Докажите, что $H(\alpha, \beta) = H(\alpha) + H(\beta \mid \alpha)$.

Информацией в случайной величине α о случайной величине β называется разница между условной и безусловной энтропиями β:

$$I(\alpha; \beta) := H(\beta) - H(\beta \mid \alpha).$$

Задача 3. Докажите, что информация симметрична, т.е., $I(\alpha; \beta) = I(\beta; \alpha)$.

В силу симметричности величину $I(\alpha; \beta)$ обычно называют *взаимной информацией* двух случайных величин.

Задача 4. Придумайте определение взаимной информации α и β при известном значении γ (предполагается, что α, β, γ имеют совместное распределение вероятностей).

Задача 5. Докажите, что для любого распределения тройки α, β, γ

$$H(\alpha, \beta, \gamma) + H(\gamma) \leq H(\alpha, \gamma) + H(\beta, \gamma).$$

Задача 6. Докажите, что для любого распределения тройки α, β, γ

$$2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\beta, \gamma) + H(\gamma, \alpha).$$

Рассмотрим на множестве $\{0,1\}^n$ два распределения вероятностей. Распределение вероятностей μ соответствует подбрасыванию n симметричных монет: вероятность каждого $x \in \{0,1\}^n$ равна $1/2^n$. Распределение вероятностей ν соответствует подбрасыванию “кривой монеты”, которая выпадает орлом с вероятностью $p > 1/2$ и решкой с вероятностью $q = 1-p$. Для каждого $x \in \{0,1\}^n$, в котором m единиц и $l = n-m$ нулей

$$\mathrm{Prob}_\nu[x] = p^m \cdot q^l.$$

Задача 7. (а) Назовем $A \subset \{0,1\}^n$ множество всех таких x, в которых нулей больше pn. Докажите, что

$$\frac{\mathrm{Prob}_\nu[A]}{\mathrm{Prob}_\mu[A]} \geq (2p^p q^q)^n.$$

(б) Пусть S состоит из всех $x \in \{0,1\}^n$, в которых единиц больше $(\frac{1}{2}+\delta)n$. Докажите, что

$$\mathrm{Prob}_\mu[S] \leq 2^{(h(\frac{1}{2}+\delta)-1)n},$$

где $h(p) = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}$.

(в)* Обозначим x результат n подбрасываний монеты, которая выпадает орлом с вероятностью p и решкой с вероятностью $q = 1-p$. Докажите, что для достаточно малых δ

$$\mathrm{Prob}_\nu[\text{число единиц в случайном } x \text{ отличается от } pn \text{ больше, чем на } \delta n] \leq 2 \cdot 2^{-cn}$$

для некоторого $c = c(p,\delta) > 0$. *Указание:* Рассмотрите “смещённые” распределения ν_1 и ν_2, в которых орёл выпадает с вероятностями $p+\delta$ или $p-\delta$ соответственно. Оцените разницу меры ν (с вероятностью выпадения орла p) и смещённых мер ν_1 и ν_2 для тех x, в которых слишком много или слишком мало единиц. Константа c окажется дивергенцией Кульбака–Лейблера для некоторой пары распределений.