Boris Kudryashov

ITMO University

October 22, 2016

1 Prove that the number of all different subsets of an n-element discrete set is equal to 2^n.

1 Prove that the number of all different subsets of an n-element discrete set is equal to 2^n.

4 Prove that $|XY| = |X||Y|$.

5 How many different sequences of length n of elements $X = \{0, ..., q - 1\}$ exist?

6 Prove the following properties of moments of random variables:

- $M[x + y] = M[x] + M[y]$.
- $M[cx] = cM[x]$.
- $M[xy] = M[x]M[y]$.
- $D[x + y] = D[x] + D[y]$.
- $D[cx] = c^2 D[x]$.
- $D[c + x] = D[x]$.
- If x and y are independent, then $K(x, y) = 0$. That is, independent random variables are uncorrelated (but not vice versa).

7 For a Discrete Memory-less Source (DML)
$X = \{a, b, c\}$ with probability distributions
- $p(a) = p(b) = p(c) = 1/3$;
- $p(a) = p(b) = 1/4$, $p(c) = 1/2$;

calculate the amount of information for each of letters and the entropy of X. Compute the amount of information in the sequence *abaac*.

8 Prove, that linear combination of convex functions with positive coefficients is also a convex function.

8 Prove, that linear combination of convex functions with positive coefficients is also a convex function.

9 Prove that

$$H\left(\frac{\boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3}{3}\right) \geq \frac{H(\boldsymbol{p}_1)}{3} + \frac{H(\boldsymbol{p}_2)}{3} + \frac{H(\boldsymbol{p}_3)}{3}$$

10 Prove theorem (1):

Theorem

Let $f(\boldsymbol{x})$ be the convex $\cap$ function of a vector argument $\boldsymbol{x}$, defined on a convex set R and $\alpha_1, ..., \alpha_M \in [0, 1]$, $\sum_{m=1}^{M} \alpha_m = 1$. Then for all $\boldsymbol{x}_1, ..., \boldsymbol{x}_M \in R$:

$$f\left(\sum_{m=1}^{M} \alpha_m \boldsymbol{x}_m\right) \geq \sum_{m=1}^{M} \alpha_m f(\boldsymbol{x}_m). \qquad (1)$$

11 Let $X = \{a, b, c\}$. Assume that $p(a) = \alpha$ is known. Draw a plot for entropy $H(X)$ as a function of $p(b) = \beta$.

11 Let $X = \{a, b, c\}$. Assume that $p(a) = \alpha$ is known. Draw a plot for entropy $H(X)$ as a function of $p(b) = \beta$.

12 Formulate conditions of α and β such that the Markov chain with the following probability transition matrix:

$$\Pi = \left[\begin{array}{cc} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{array} \right]$$

will generate the sequence of independent random symbols?

13 Without performing calculations, write down stationary probability distribution for the Markov chain with the probability transition matrix:

$$\Pi = \begin{bmatrix} 1/2 & 1/2 \\ 0 & 1 \end{bmatrix}.$$

14 *Non-ergodic Markov chain* Find the stationary
probability distribution for the Markov chain
with the probability transition matrix:

$$\begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 3/4 & 0 & 1/4 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix}.$$

15 Binary source is defined as the Markov chain with the probability transition matrix:

$$\Pi = \left[\begin{array}{cc} 1/2 & 1/2 \\ 1/8 & 7/8 \end{array} \right].$$

Find the amount of information in the sequence 0110. Compute $H(X)$. Derive formulas $H(X|X^n)$ and $H_n(X)$ as functions of n, find $H(X|X^\infty)$. What impact has source memory on the information rate of the source?

16 *Representation of binary sequences in a form of series-length sequence.* A binary sequence, containing a small number of ones, can be compactly written as a sequence of lengths of series of zeros. Denote 0^i a sequence of i zeros. As example, the sequence 0011000010100 will be written as (2,0,4,1,2) in this notation. Thus, binary sequence $x_1, ..., x_n, ..., x_i \in X = \{0, 1\}$ will be transformed into the sequence of non-negative integer numbers – $y_1, y_2, ..., y_i \in Y = \{0, 1, ...\}$, where y_i are lengths of series of zeros. Consider BMS with $p(\text{"1"}) = p$, find the probability distribution for the lengths of series Y, entropy $H(Y)$.

16 *Representation of binary sequences in a form of series-length sequence.* A binary sequence, containing a small number of ones, can be compactly written as a sequence of lengths of series of zeros. Denote 0^i a sequence of i zeros. As example, the sequence 0011000010100 will be written as (2,0,4,1,2) in this notation. Thus, binary sequence $x_1, ..., x_n, ..., x_i \in X = \{0, 1\}$ will be transformed into the sequence of non-negative integer numbers – $y_1, y_2, \ldots, y_i \in Y = \{0, 1, ...\}$, where y_i are lengths of series of zeros. Consider BMS with $p(\text{"1"}) = p$, find the probability distribution for the lengths of series Y, entropy $H(Y)$.
Hint: While solving the problem, you will get the geometric probability distribution, and while entropy calculations, you will need the mathematical expectation formula. See the next problem.

17 *Geometric probability distribution* on infinite set
$I = \{0, 1, 2, ...\}$ is defined as:
$p(i) = (1 - \alpha)\alpha^i$
Why is it called so? Check the probability
normalization condition. Find mathematical
expectation, variation and entropy $H(I)$.

17 Hint: To check probability normalization, you calculate the sum:

$$\sum_{i=0}^{\infty} a^i. \tag{2}$$

Use geometric progression sum formula. To calculate the mathematical expectation, you should deal with the sum:

$$\sum_{i=0}^{\infty} i a^i. \tag{3}$$

Note, that derivative of (2) by a, gives almost (3). Thus, we have reduced the problem to the differentiation of geometric progression sum formulae. The second derivative should be used for variation calculation. One more way for calculating sum (3) is, that the factor i under the sum sign can be represented as the sum i ones, and then the order of the summation is changed.
Answers:

$$\mathbf{M}[i] = \frac{\alpha}{1-\alpha}; \quad \mathbf{D}[i] = \frac{\alpha}{(1-\alpha)^2}; \quad H(I) = \frac{\eta(\alpha)}{1-\alpha}.$$

18 *The relative entropy* of distribution $\{p(x), x \in X\}$ with respect to distribution $\{q(x), x \in X\}$ is defined by:

$$L(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}. \tag{4}$$

This function measures the difference of two distributions. Prove, that the relative entropy is equal to zero, if this two distributions are equal, and it is strictly positive if they differ.

18 *The relative entropy* of distribution $\{p(x), x \in X\}$ with respect to distribution $\{q(x), x \in X\}$ is defined by:

$$L(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}. \qquad (4)$$

This function measures the difference of two distributions. Prove, that the relative entropy is equal to zero, if this two distributions are equal, and it is strictly positive if they differ.

Hint. Rewrite the relative entropy like:

$$L(p||q) = -\sum_{x \in X} p(x) \log \frac{q(x)}{p(x)}.$$

and use the inequality $\log z \leq (z - 1) \log e$.

19 Uniform distribution is entropy-maximizing in the sense that it has the largest entropy among all probability distributions on the set of finite fixed size.

Prove that a *Geometric distribution is entropy-maximizing for discrete random variables over countable alphabets having fixed mathematical expectation.*

19 *Solution.* Let $p(i) = (1-\alpha)\alpha^i, i = 0, 1, ...$ be a geometric distribution, and $q(i)$ be an arbitrary distribution on the same set of numbers with the same mathematical expectation.
$\mathbf{M}[i] = \alpha/(1-\alpha)$, $H(p)$ и $H(q)$ – corresponding entropy's, such that $H(p) = \eta(\alpha)/(1-\alpha)$ (see the 17-th pronlem). Then,

$$
\begin{aligned}
H(q) &= -\sum_{i=0}^{\infty} q(i) \log q(i) = -\sum_{i=0}^{\infty} q(i) \log\left(\frac{q(i)}{p(i)}p(i)\right) = \\
&= -L(p\|q) - \sum_{i=0}^{\infty} q(i) \log p(i) \leq \\
&\leq -\sum_{i=0}^{\infty} q(i)\left(\log(1-\alpha) + i \log \alpha\right) = \\
&= -\log(1-\alpha) - \mathbf{M}[i] \log \alpha = \\
&= -\log(1-\alpha) - \frac{\alpha \log \alpha}{1-\alpha} = H(p).
\end{aligned}
$$

20 Consider DMS over $X = \{a, b, c\}$. For probability distributions,

    ❶ $p(a) = p(b) = p(c) = 1/3$;
    ❷ $p(a) = p(b) = 1/4$, $p(c) = 1/2$;

investigate the dependency between *uniform code rate* and length of encoded blocks for the case of error-free coding. Compare the code rate with the entropy of the source.

20 Consider DMS over $X = \{a, b, c\}$. For probability distributions,

    ❶ $p(a) = p(b) = p(c) = 1/3$;
    ❷ $p(a) = p(b) = 1/4$, $p(c) = 1/2$;

investigate the dependency between *uniform code rate* and length of encoded blocks for the case of error-free coding. Compare the code rate with the entropy of the source.

21 Consider DMS from the previous problem. Investigate the dependency between code rate of uniform code and length of encoded sequences with the error probability equal to 0.1.

22 Hamming weight of sequences $\boldsymbol{x}$ over the alphabet $\{0, \ldots, q-1\}$ is defined as the number of non-zero elements in $\boldsymbol{x}$. How many different binary sequences of length n and w exist?

- *Hint:* There are choices of w elements from the set of n elements (the order pf choosing is taken into account):

$$A_n^w = n(n-1) \times \cdots \times (n-w+1).$$

There are P_w different permutations of w elements.

$$P_w = w(w-1) \times \cdots \times 1 = w!.$$

Thus, the number of sequences of length n and weight w is equal to the number of *Combinations* of w elements from n elements ignoring the order of elements, and is equal to:

$$C_n^w = \binom{n}{w} = \frac{A_n^w}{P_w} = \frac{n(n-1) \times \cdots \times (n-m+1)}{w!} =$$

$$= \frac{n!}{w!(n-w)!}.$$

23 Find the probability $p_n(w)$, that a sequence of length n at the output of BMS with $p(1) = p$, will have Hamming weight equal to w.

23 Find the probability $p_n(w)$, that a sequence of length n at the output of BMS with $p(1) = p$, will have Hamming weight equal to w.

The answer is the binomial distribution:

$$p_n(w) = \binom{n}{w} p^w (1-p)^{n-w}.$$

24 BMS $X = \{0, 1\}$ generates ones with probability $p = 0,02$. Consider the dependence between error probability and length of encoded blocks, with the code rate 0.5 bit for a letter of source.

24 BMS $X = \{0, 1\}$ generates ones with probability $p = 0,02$. Consider the dependence between error probability and length of encoded blocks, with the code rate 0.5 bit for a letter of source.

25 Find the probability that the sequence of length $n = 10$ at the output of BMS with $p(1) = 0,1$, will have Hamming weight greater or equal to $w = 5$. (Use your PC). Compare this result with the Chebyshev's estimate on the probability of this event.

26 *Polynomial distribution.* How many ternary sequences of length n contain τ_0, τ_1 and $\tau_2 = n - \tau_0 - \tau_1$ elements of 0, 1 and 2 respectively?
Find the probability of such a sequence given probability distribution on elements of ternary alphabet. Generalize the result to the alphabets of arbitrary size.