Problems to the mini-course
"Shannon, codes, practical computations" (RMS, Sept. 2012)

Each problem has a weight attached (an integer number in parentheses). Excellent rating is earned by solving a set of problems with total weight 20+, good rating 15+, positive rating 10+. Please send your solutions to {kolya,sergei}@bach.inm.ras.ru — using TEX or JPEGs, even better.

1. **(1)** For the function

$$H(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}, \qquad 0 < x < 1, \tag{1}$$

prove that it is (a) convex, (b) symmetric with respect to $x = \frac{1}{2}$, (c) monotonously increasing in interval $x \in (0, \frac{1}{2})$.

2. **(3)** Using the Stirling approximation, prove

$$\log \binom{N}{k} \approx NH(k/N), \qquad N \to \infty.$$

Provide the next asymptotic term, if possible. The logarithm base here is the same as in (1).

3. **(4)** Consider a code with N repetitions over a BSC(p) channel. Prove that the block error probability is

$$P_E = \sum_{\frac{N}{2} < k \leqslant N} \binom{N}{k} p^k (1-p)^{(N-k)}.$$

Also, using the Stirling approximation, find the first asymptotic term for $N \to \infty$. Take $p = 0.01$ and estimate the code length N which will guarantee $P_E < 10^{-15}$.

4. **(3)** [Calibrating weights] One has a collection of weights 1, 2, 3, 4 grams, but one of the weights is a fake. Can you use only *two* weighings to determine which one is a fake and is it lighter or heavier than prescribed?

5. **(1)** Recall the definitions of Hamming distance and Hamming weight. Prove that for a linear code with codewords c_i,

$$\min_{i,j} d_H(c_i, c_j) = \min_i w_H(c_i).$$

6. **(6)** [Concatenation of codes] Consider a code with 9 repetitions, but nontrivial encoding/decoding procedures: source bits are first encoded with 3 repetitions, and next the output (3 bit words) are again encoded with 3 repetitions. The decoder is also two-stage.

   - estimate the block error probability of the concatenated code;
   - estimate the complexity of encoding/decoding procedures — is something gained with respect to usual 9 repetitions code?

7. [Several applications for conditional probabilities]

   (a) **(2)** Young man has passed some medical test. The reliability of this test is 95%. It is also known that at this very age, only 1% of young men have this particular disease. What are the chances that the young man is ill?

   (b) **(3)** There are eleven urns numbered $0, 1, \ldots, 10$. An urn labeled k contains k black and $10 - k$ white balls, well mixed together. Fred chooses randomly an urn and draws randomly n balls, each time with replacement. After $n = 10$ draws the number of black balls observed was $n_B = 3$. What is the probability that the urn number k was chosen? What is the probability that the next ball will be black?

   (c) **(3)** Fred tosses the coin n times, getting a random sequence of heads and tails. The probability p that the coin will land heads up, is unknown. Estimate p, if the number of heads is n_H.

   (d) **(2)** [Monty Hall paradox] You are given a choice of 3 doors: behind one door is a prize; behind the others, nothing. You pick a door (but the door is not opened), and the host opens another door, which has nothing behind it. Should you stay with your original decision or should you switch your choice?

   (e) **(4)** Fewer than 1 in 1000 women who are abused by their mates go on to be killed by them. It is known that Mr. S. has beaten his wife, which was found dead. Determine the statictical probability of Mr. S. guilt, assuming something sane if you need more information (total number of women murders per year, lifetime of a marriage with a beating husband, percent of murders committed by husband, total number of women living etc.)

8. **(5)** There are 3 cards with sides colored white/white, white/black, black/black. We shuffle the cards (meaning that the card orientation is also random) and draw one card. The upper side happens to be black. What will be the color of the lower side? In answering this question statistically, is it important to know the color of the upper side? Calculate the entropy and information for the result of each experiment. Let u be the color of the upper side and l the color of the lower side. Find $H(U)$, $H(L)$, $I(U; L)$.

9. **(1)** [Application of Jensen inequality] Given two vectors $p = [p_1, \ldots, p_n]$ and $q = [q_1, \ldots, q_n]$, define $A(p, q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$ (which is an example of *Bregman distance*). Prove that $A(p, q) \geqslant 0$ under the conditions $p_i \geqslant 0$ and $\sum_i p_i = \sum_i q_i$. Moreover, prove that $A = 0$ is equivalent to $(p_i - q_i)p_i = 0$ for all $i = 1, \ldots, n$.

10. **(1)** Prove that $I(X; Y) = I(Y; X)$.

11. **(3)** Express the mutual information I through the joint probabilities of X and Y,

$$I(X, Y) = \sum_{x \in A_X} \sum_{y \in A_y} P(x, y) \log \left(\frac{P(x, y)}{P(x)P(y)} \right)$$

    Was the previous problem helpful here?

12. **(3)** [Compression of information] Consider a discrete memoryless channel $X \to Y$. Join elements of A_Y in (arbitrary) non-intersecting subsets z_1, z_2, ... which will be values of new random variable Z: event $\{Z = z_j\}$ is by definition $\cup_{y_i \in z_j}\{Y = y_i\}$. Define the probability function p_Z using p_Y and prove

$$I(X;Z) = I(X;Y) - \sum_y p_Y(y) A\big(p(x|y), p(x|z(y))\big),$$

where arguments of the function A are vectors of length $|A_X|$ (with y fixed). Deduce that lossy compression of output signal Y can be achieved by choosing Z (given $|A_Z|$) so that Bregman distance between Y and Z is minimized.

13. **(2)** How many elements of $\mathbb{F}_2^n$ have Hamming distance to origin not exceeding t? Estimate this *volume of Hamming ball of radius* t in closed form.

14. **(3)** Show that for BSC(p) with $p < \frac{1}{2}$, maximum likelihood (ML) decoder and minimum distance decoder are equivalent. Is it possible to modify minimum distance decoder so that it will work like ML for $p > \frac{1}{2}$?

15. **(2)** *Extended* Hamming code has codewords of plain Hamming code extended by a parity check bit: the word $c = (c_1 c_2 \dots c_n)$ is extended to $(c_1 c_2 \dots c_n \oplus_{i=1}^n c_i)$. Let (10100111) and (10011111) be noisy codewords of extended Hamming code with 1 and 2 bits flipped. Which one has 2 bits flipped?

16. **(2)** Using the code with check matrix

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix},$$

encode the messages (0110) and (1010). Also, decode noisy words (1100000), (1001010), (1101011), (0110110), (0100111). How many errors have you detected?

17. **(4)** Starting from factor graph approach, derive the decoding algorithm for the BEC with as much simplifications as possible. What is the message alphabet and what are computation rules?

18. **(5)** [Hänsel und Gretel verliefen sich im Wald] Hänsel and Gretel, together with all their classmates, went for a walk in the woods. The forest is so dark that each kid can only see its immediate neighbors. Assume that communication is limited to these nearest neighbors as well and that all the kids form a tree (in the graph sense) with respect to this neighborhood structure.

    Construct a message-passing algorithm which allows them to count, to ensure that none of the children was eaten by the wolf. What is the initialization? The message-passing rules? Can you modify the algorithm to only count a prescribed subset, e.g. only girls?

19. **(6)** Let all circle nodes in Tanner graph have degree 2, and let all square nodes have degree 3. Let the code length n be fixed. Consider a graph ensemble subject to this distribution. Note that all circle nodes are statistically equivalent and starting from some circle, build a random tree corresponding to one BP iteration.

- which graph algorithms are possible?
- what is the probability measure corresponding to a graph algorithm?
- what are the erasure probabilities for each possible graph algorithm?
- explain the work of BP on this example.