# Statistical Information Theory MT7037 VT20
# Take home project 1

Examiner: Chun-Biu Li (cbli@math.su.se)
Posting date: Feb 7, 2020

---

***** Read carefully the following instructions before starting *****

- Deadline: **8pm Feb 17, 2020**
- The submission should include your report to the questions (in .pdf file) and the signed confirmation sheet. The files should be submitted at the course page.
- Total points is 40 in this project. A 6-point deduction will be imposed for late submission.
- Only the reports of LADOK registered students, except for PhD students, will be graded.
- Credit points will be given based on clear logical explanation and steps leading to the final solution. Also redundant writing irrelevant to the solution can result in point deduction.
- The report, not including the appendices, should be $\leq 6$ pages.
- The report must be completed independently. Plagiarism or other forms of cheating is a serious act. To underline this, the *signed confirmation sheet* must be submitted to declare that your work is made in accordance with the rules for written exams at Stockholm University (see course page).

---

## Exercise 1 (Total: 17p)

This exercise covers some basic concepts of information theory discussed in the classes. Hint: This exercises can be completed with few-step derivations!

a) Let $X$ and $Y$ be discrete random variables, show that $H((X+Y)|X) = H(Y|X)$. Then show that $H(X+Y) \geq H(Y)$ if $X$ and $Y$ are independent. (5p)

b) In the course book (Cover & Thomas 2nd Ed.), sufficient statistic is defined in terms of the mutual information (Eq. 2.124 in Cover & Thomas). Show that this information-theoretic definition of sufficient statistic is the same as the one defined in your previous statistical courses. (5p)

c) Let $\mathbf{X}$ be an $n$-dimensional, continuous random variable, show that the relative entropy $D(f_{\mathbf{X}}(\mathbf{x}) \| h_{\mathbf{X}}(\mathbf{x}))$ is invariant under a bijective transformation, $\mathbf{Y} = g(\mathbf{X})$, from $\mathbf{X}$ to $\mathbf{Y}$. (3p)

d) In class, we discussed the decomposition of the rate of Shannon entropy change for continuous time Markov chain as follows: $dS/dt = dS_i/dt + dS_e/dt$, where $S = -\sum_i p_i \ln p_i$ is the Shannon entropy of the system, $dS_i/dt = \frac{1}{2}\sum_{i,j} J_{ij} \ln \frac{\langle i|j\rangle p_j}{\langle j|i\rangle p_i}$ denotes the rate of entropy change due to *irreversible* dynamics, and $dS_e/dt = -\frac{1}{2}\sum_{i,j} J_{ij} \ln \frac{\langle i|j\rangle}{\langle j|i\rangle}$ denotes the rate of entropy flow from the system to the environment. Here $p_i := p(x_t = i)$, $\langle i|j\rangle := p(x_{t+dt} = i|x_t = j)/dt$, $J_{ij} := \langle i|j\rangle p_j - \langle j|i\rangle p_i$ is the "net" probability flow from state $i$ to $j$. Show that $dS_i/dt \geq 0$ (i.e., the 2nd law of thermodynamics) and state when the equality holds. (4p)

## Exercise 2 (Total: 23p)

After learning the basic notions of transfer entropy, this exercise allows you to think critically if the transfer entropy really carries the interpretation of information flow. To complete this exercise, you need to first download and read the article by James *et al.*, "Information flows? A critique of transfer entropy", *Phys. Rev. Lett.*, vol. 116, p238701 (2016). **Note:** Notations in this exercise are the same as those in the paper.

Consider the first example, *two time series*, in page 2 of the paper:

a) Verify that $\mathcal{T}_{X\to Y} = 1$ bit. (5p)

b) Verify that $\mathcal{I}[Y_t : (X_{0:t}, Y_{0:t})] = 1$ bit. (5p)

c) Verify that the coinformation, $\mathcal{I}(Y_t : X_{0:t} : Y_{0:t}) = -1$ bit (see left column of page 3 of the paper). Hint: Derivation can be done in a few lines! (4p) *Take home message*: This is an example showing that the area-entropy correspondence in the Venn diagram breaks down in $n$-variable (n>2) case.

d) Show that the coinformation can also be expressed as $\mathcal{I}(Y_t : X_{0:t} : Y_{0:t}) = [\mathcal{I}(Y_t : X_{0:t}) + \mathcal{I}(Y_t : Y_{0:t})] - \mathcal{I}[Y_t : (X_{0:t}, Y_{0:t})]$. Hint: Derivation can also be done in a few lines! (4p) *Take home message*: The coinformation has been used to quantify synergy (when $\mathcal{I}(Y_t : X_{0:t} : Y_{0:t}) < 0$) and redundancy (when $\mathcal{I}(Y_t : X_{0:t} : Y_{0:t}) > 0$) of using two variables to predict a third variable.

Now consider the second example, *three time series*, in page 3 of the paper:

e) Verify that $C_{X\to Z|(Y,Z)} = C_{Y\to Z|(X,Z)} = 1$ bit. Note: These correspond to the conditional transfer entropy in the book "An introduction to transfer entropy" covered in the class. (5p)

**Final remark**: Although concepts in the "Discussion" and "Conclusion" sections of the paper were not asked in this exercise, the modern developments in information theory, such as the partial information decomposition, unique information, hypergraph, $\epsilon$-machine, information thermodynamics, etc., mentioned in these sections are interesting ideas to be explored in the future. In particular, $\epsilon$-machine, which is the most minimal but most predictive nonparametric time series model, will be covered in the later part of the course.

*Good Luck!*