STAT210A Theoretical Statistics	Lecture 0 - 01/20/2021

Lecture 0: Lecture scribing template

Instructor: Song Mei *Scriber: Song Mei* *Proof reader: Song Mei*

1 Conventions

I typically use the following conventions in my own notes. These conventions are suggested to the scribers but not strictly enforced. The scriber can focus on scribing the notes while keeping in mind these conventions. The proof reader's job is to ensure the consistency of notations across different lecture notes.

- Use $\mathbb{R}$, $\mathbb{C}$, $\mathbb{N}$, $\mathbb{S}$, $\mathbb{Z}$, $\{-1, 1\}$ for sets. Use boldface uppercase letters for matrices, boldface lowercase letters for vectors, no boldface lower letters for scalers.
  For example: $i, j \in \mathbb{N}$, $a, b \in \mathbb{R}$, $z, \xi \in \mathbb{C}$, $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{A} \in \mathbb{R}^{n\times d}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$, $\boldsymbol{\mu}, \boldsymbol{\sigma} \in \{-1, 1\}^n$, $\boldsymbol{v} \in \mathbb{S}^{d-1}(r)$ where r is the radius of the d-dimensional sphere. Moreover
  $$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{\varepsilon} \tag{1}$$
  where $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{A} \in \mathbb{R}^{n\times d}$, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\boldsymbol{\varepsilon} \in \mathbb{R}^n$.
- Vectors are always column vectors. Use "\sT" T for vector/matrix transpose. When using subscripts to refer to elements of a matrix or a vector, remove the boldface.
  For example: $\boldsymbol{y} = (y_1, \ldots, y_n)^\mathsf{T} \in \mathbb{R}^n$, $\boldsymbol{x}_0 = (x_{0,1}, \ldots, x_{0,d})^\mathsf{T} \in \mathbb{R}^d$.
- Use $\mathbb{P}$ for probability. Use $\mathbb{E}$ for expectation. When there are multiple independent random variables and you want to look at conditional expectation with respect to a subset of the random variable, or when you want to emphasize the distribution of the random variable that you take expectation with respect to, use subscript $\mathbb{P}_{G\sim\mathcal{N}(0,1)}$, $\mathbb{E}_{G\sim\mathcal{N}(0,1)}$. Notation $\mathcal{N}(0, 1)$ is for standard Gaussian. Notation $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ is for multi-variates isotropic Gaussian. For example:
  $$\begin{aligned}\mathbb{P}(X^2 \geq t^2) &\leq \frac{\mathbb{E}[X^2]}{t^2}, \\ \mathbb{E}_{G\sim\mathcal{N}(0,1)}[(X + G)^2] &\equiv \mathbb{E}[(X + G)^2|X] = X^2 + 1.\end{aligned} \tag{2}$$
- Use $\sim_{i.i.d.}$ for identically distributed as. For example: $\varepsilon_i \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$.
- Avoid redefining the left and right parenthesis using a single command. Literally use "\vert" for absolute value $|a|$ and "\Vert" for norm $\|\boldsymbol{A}\|_{\rm op}$. Try to avoid using "\left" and "\right" to adjust the size of parenthesis and norm symbols. Use "\Big" (more frequently) and "\big" (less frequently) instead. For example
  $$\mathbb{P}\Big(\Big\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\Big\|_{\rm op} \geq t\Big) \leq c_1 d e^{-c_2 n t^2/d}.$$
- Avoid using "\frac" in inline equations, i.e., avoid using $\frac{1}{n}\sum_{i=1}^n x_i$ in inline equations. Using $(1/n)\sum_{i=1}^n x_i$ or $n^{-1}\sum_{i=1}^n x_i$ instead. You can freely use "\frac" in out of line equations
  $$\bar{x} \equiv \frac{1}{n}\sum_{i=1}^n x_i.$$

- Use $X_d = o_d(1)$ to denote $\lim_{d\to\infty} X_d = 0$. Use $X_d = O_d(1)$ to denote $\limsup_{d\to\infty} X_d < \infty$. Use $X_d = o_{d,\mathbb{P}}(1)$ to denote X_d converges to 0 in probability. Use $X_d = O_{d,\mathbb{P}}(1)$ to denote X_d is bounded with probability goes to 1 as $d \to \infty$.

- For self-defined symbols, add their definitions at the beginning of the tex file (don't add them in the sty file). For example, I defined a symbol GOE for this lecture notes. You can find it at the beginning of the tex file. Here I use this symbol in the lecture notes $\boldsymbol{W} \sim \mathrm{GOE}(n)$.

- Use "\[" and "\]" for single line no number equations. Use "\[\begin{aligned}" and "\] \end{aligned}" for multiline no number equations. Use "\begin{equation}\label{ }" and "\end{equation}" for single line numbered equations with labels. Use "\begin{equation}\label{} \begin{aligned}" and "\end{aligned} \end{equation}" for multiline numbered equations with labels. Use "&~" and "\\" to align and separate different lines. For example, we have

$$e^{i\pi} + 1 = 0, \tag{3}$$

and

$$\begin{aligned} \hat{\boldsymbol{v}} =&~ \arg\max_{\boldsymbol{v}\in\mathbb{S}^{d-1}} \langle \boldsymbol{v}, \boldsymbol{A}\boldsymbol{v}\rangle \\ =&~ \boldsymbol{v}_{\max}(\boldsymbol{A}). \end{aligned} \tag{4}$$

If you want to give different labels to different lines of equations, use "\begin{align}" and "\end{align}". For example

$$-\partial_\beta \Phi_n(\beta,\lambda) = \langle H_n\rangle_{\beta,\lambda}, \tag{5}$$

$$-\frac{1}{\beta}\partial_\lambda \Phi_n(\beta,\lambda) = \langle f_n\rangle_{\beta,\lambda}, \tag{6}$$

- Use labels to refer to equations, lemmas, theorems, definitions, remarks, corollaries, propositions, and formalisms. Label equations starting with "eqn:". Label definitions starting with "def:". Label lemmas starting with "lem:". Label theorems starting with "thm:". Label remarks starting with "rmk:". Label corollaries starting with "cor:". Label propositions starting with "prop:". Label formalisms starting with "form:". For example: Recall the definition of $\boldsymbol{y}$ as in Eq. (1). By Lemma 2 below, we have $\Phi_n''(\beta,\lambda) \ge 0$.

- Add a reference using "\cite" command: [BM11]. Add bibliography using "\bibliography" command (see the end of the tex document).

2 Environments

We use the definition environment for definitions.

Definition 1 (Free entropy)**.** *Let $H_n, f_n : \Omega \to \mathbb{R}$ be two measurable functions, and $\beta \ge 0$ and $\lambda \in \mathbb{R}$. We define the free entropy function $\Phi_n : \Sigma \to \mathbb{R}$ associated to the Hamiltonian $H_n + \lambda f_n$ by*

$$\Phi_n(\beta,\lambda) = \log \int_\Omega \exp\Big\{ -\beta\Big(H_n(\boldsymbol{\sigma}) + \lambda f_n(\boldsymbol{\sigma})\Big)\Big\}\nu_0(\mathrm{d}\boldsymbol{\sigma}).$$

The domain of $\Sigma \subseteq \mathbb{R}^2$ is such that Φ_n is well-defined (integration exists).

We use lemma/proposition/theorem environments for rigorous results.

Lemma 2 (Free entropy is convex in the inverse temperature)**.** *Assume Φ_n exists and is second-times differentiable. Then for any $\beta > 0$ and $\lambda \in \mathbb{R}$, we have*

$$\Phi_n''(\beta,\lambda) \ge 0.$$

Proof of Lemma 2. Recall the definition of free entropy $\Phi_n(\beta, \lambda)$ as in Definition 1 (Use labels to refer to definitions, lemmas, theorems, and propositions. Please avoid to use fixed numbers). We have

$$\Phi_n''(\beta, \lambda) = \mathrm{Var}_{\beta,\lambda}((H_n + \lambda f_n)) \geq 0.$$

This concludes the proof. $\square$

Theorem 3 (Theorem name, or references)**.** *Theorems are more important than Propositions, and Propositions are more important than lemmas.*

In this course, we will make many important non-rigorous statements. If an important statement is non-rigorous, we use formalism environment instead of lemma/proposition/theorem. For example, we have the following formalism

Formalism 4 (Formalisms are non-rigorous lemmas/propositions/theorems)**.** *Let $\Phi_n(\beta, 0)$ be as defined in Definition 1. Under certain conditions on the sequence H_n and f_n, we have*

$$\lim_{\beta\to\infty} \lim_{n\to\infty} \frac{1}{\beta n} \Phi_n(\beta, 0) = - \lim_{n\to\infty} \min_{\boldsymbol{\sigma}} H_n(\boldsymbol{\sigma})/n. \tag{7}$$

We use remark to make comments.

Remark 5. *The formalism above assumed that $\lim_{n\to\infty}$ and $\lim_{\beta\to\infty}$ can be exchanged. For certain H_n and f_n, this can be proved in a case by case analysis.*

References

[BM11] Mohsen Bayati and Andrea Montanari, *The lasso risk for gaussian matrices*, IEEE Transactions on Information Theory **58** (2011), no. 4, 1997–2017.