STAT260 Mean Field Asymptotics in Statistical Learning

Concentration inequalities

Lecturer: Song Mei *Assignment 1 - Due on 2/14/2021*

Homework submissions are expected to be in pdf format produced by LaTeX. For coding exercises, you are free to choose your favorite programming language, but matlab or Python are highly recommended. Grading of coding exercises will base on the figures produced by the simulations, but you need also submit the source code in the original format (e.g. '.m' or '.py').

Coding exercises

Exercise 1: Eigenvalue and eigenvectors of the spiked matrix model

Let $\boldsymbol{v} \in \mathbb{S}^{n-1} \equiv \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_2^2 = 1\}$ be any fixed vector. We observe a matrix $\boldsymbol{Y} = \lambda \boldsymbol{v}\boldsymbol{v}^\mathsf{T} + \boldsymbol{W}$, where $\boldsymbol{W} \sim \mathrm{GOE}(n)$. That is, $\boldsymbol{W} \in \mathbb{R}^{n\times n}$, $W_{ij} \sim_{iid} \mathcal{N}(0, 1/n)$ for $i < j$, and $W_{ii} \sim_{iid} \mathcal{N}(0, 2/n)$, and $\boldsymbol{W} = \boldsymbol{W}^\mathsf{T}$. The parameter λ is the signal-to-noise ratio and is a fixed real number that does not grow with n. Given observation $\boldsymbol{Y}$, we would like to estimate $\boldsymbol{v}$. The maximum likelihood estimator gives the spectral estimator $\hat{\boldsymbol{v}}(\boldsymbol{Y})$:

$$\hat{\boldsymbol{v}}(\boldsymbol{Y}) = \arg\max_{\boldsymbol{u}\in\mathbb{S}^{n-1}} \langle \boldsymbol{u}, \boldsymbol{Y}\boldsymbol{u}\rangle.$$

We further denote

$$\lambda_{\max}(\boldsymbol{Y}) = \max_{\boldsymbol{u}\in\mathbb{S}^{n-1}} \langle \boldsymbol{u}, \boldsymbol{Y}\boldsymbol{u}\rangle = \langle \hat{\boldsymbol{v}}(\boldsymbol{Y}), \boldsymbol{Y}\hat{\boldsymbol{v}}(\boldsymbol{Y})\rangle, \qquad u(\boldsymbol{Y}) = \langle \hat{\boldsymbol{v}}(\boldsymbol{Y}), \boldsymbol{v}\rangle^2.$$

We have shown in class that $\lambda_{\max}(\boldsymbol{Y})$ (and actually, also $u(\boldsymbol{Y})$) will concentrate well around their expectation, i.e., we have

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big|\lambda_{\max}(\boldsymbol{Y}) - \mathbb{E}[\lambda_{\max}(\boldsymbol{Y})]\Big| \ge \varepsilon\Big) = 0, \qquad \lim_{n\to\infty} \mathbb{P}\Big(\Big|u(\boldsymbol{Y}) - \mathbb{E}[u(\boldsymbol{Y})]\Big| \ge \varepsilon\Big) = 0.$$

In this exercise, we perform some simulations to show how well they concentrate for finite n.

Please do the following simulation: for each $n = 50, 200, 800$, and each $\lambda \in [0, 3]$ with grid size 0.1, obtain $n_s = 20$ independent samples of matrices $\boldsymbol{Y}^{(i)} \sim_{iid} \boldsymbol{Y}$, $i \in [n_s]$. Compute the sample mean and sample standard deviation of $\{\lambda_{\max}(\boldsymbol{Y}^{(i)})\}_{i\in[n_s]}$ and $\{u(\boldsymbol{Y}^{(i)})\}_{i\in[n_s]}$, i.e., for $f \in \{\lambda_{\max}, u\}$,

$$\hat{\mathbb{E}}[f(\boldsymbol{Y})] \equiv \frac{1}{n_s}\sum_{i=1}^{n_s} f(\boldsymbol{Y}^{(i)}), \qquad \widehat{\mathrm{std}}[f(\boldsymbol{Y})]^2 \equiv \frac{1}{n_s}\sum_{i=1}^{n_s}\Big(f(\boldsymbol{Y}^{(i)}) - \hat{\mathbb{E}}[f(\boldsymbol{Y})]\Big)^2.$$

For each $f \in \{\lambda_{\max}, u\}$, make an error bar plot of $(\hat{\mathbb{E}}[f(\boldsymbol{Y})], \widehat{\mathrm{std}}[f(\boldsymbol{Y})])$ versus λ (λ will be the x-axis), with $\hat{\mathbb{E}}[f(\boldsymbol{Y})]$ the mean variable and $\widehat{\mathrm{std}}[f(\boldsymbol{Y})]$ the error variable (cf. `https://www.mathworks.com/help/matlab/ref/errorbar.html`). Please plot the curves for $n = 50, 200, 800$ in the same figure. So in total there will be two figures, one for $f = \lambda_{\max}$ and one for $f = u$.

We will show in later lectures that

$$\lim_{n\to\infty} \mathbb{E}[\lambda_{\max}(\boldsymbol{Y})] = \begin{cases} 2, & \lambda \le 1, \\ \lambda + 1/\lambda, & \lambda > 1. \end{cases} \qquad \lim_{n\to\infty} \mathbb{E}[u(\boldsymbol{Y})] = \begin{cases} 0, & \lambda \le 1, \\ 1 - 1/\lambda^2, & \lambda > 1. \end{cases}$$

Please also plot these theoretical predictions in the same figure.

Exercise 2: The LASSO risk

In this exercise, we confirm the predictions in the following theorem.

Theorem 1 (Bayati and Montanari, 2011)**.** *We consider the asymptotic limit when $n/d \to \delta \in (0, \infty)$ as $d \to \infty$. Let $\boldsymbol{A} \in \mathbb{R}^{n \times d}$ with $A_{ij} \sim \mathcal{N}(0, 1/n)$. Let $\boldsymbol{x}_0 \in \mathbb{R}^d$ with $\boldsymbol{x}_{0,i} \sim_{iid} \mathbb{P}_0$, $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \mathbf{I}_n)$, and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{w}$. Let $\hat{\boldsymbol{x}}$ be the LASSO estimator*

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2n} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \frac{\lambda}{n} \|\boldsymbol{x}\|_1. \tag{1}$$

Then we have

$$\lim_{d,n \to \infty, n/d \to \delta} \frac{1}{d} \|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2^2 = \mathbb{E}_{(X_0, Z) \sim \mathbb{P}_0 \times \mathcal{N}(0,1)} [(\eta(X_0 + \tau_\star Z; \theta_\star) - X_0)^2],$$

where $\eta(x;\theta) = sign(x) \cdot (|x| - \theta)_+$ is the soft thresholding function and $\tau_\star = \tau_\star(\alpha_\star)$. Here, define $\tau_\star(\alpha)$ to be the largest solution of

$$\tau^2 = \sigma^2 + \delta^{-1} \mathbb{E}_{(X_0, Z) \sim \mathbb{P}_0 \times \mathcal{N}(0,1)} \{[\eta(X_0 + \tau Z; \alpha\tau) - X_0]^2\},$$

and $\alpha_\star$ to be the unique non-negative solution of

$$\lambda = \alpha \tau_\star(\alpha) \cdot \Big[1 - \delta^{-1} \mathbb{E}[\eta'(X_0 + \tau_\star(\alpha) Z; \alpha \tau_\star(\alpha))]\Big].$$

Moreover, for any Lipschitz function ψ, we have almost surely

$$\lim_{d \to \infty} \frac{1}{d} \sum_{i=1}^{d} \psi(\hat{x}_i, x_{0,i}) = \mathbb{E}_{(X_0, Z) \sim \mathbb{P}_0 \times \mathcal{N}(0,1)} [\psi(\eta(X_0 + \tau_\star Z; \alpha_\star \tau_\star), X_0)].$$

Let us fix the model parameters to be the following

$$\begin{aligned} &\delta = n/d = 0.64, \\ &\sigma^2 = 0.2, \\ &\mathbb{P}_0(x = 1) = \mathbb{P}_0(x = -1) = 0.064, \qquad \mathbb{P}_0(x = 0) = 1 - \mathbb{P}(x = 1) - \mathbb{P}(x = -1). \end{aligned}$$

Question 1

Please do the following simulation: for each $d = 50, 200, 800$ (recall $n = \delta d$), and each $\lambda \in [0, 3]$ with grid size 0.1, obtain $n_s = 10$ **independent** samples of instances $(\boldsymbol{A}^{(i)}, \boldsymbol{x}_0^{(i)}, \boldsymbol{y}^{(i)})$, $i \in [n_s]$. Compute the LASSO estimator $\hat{\boldsymbol{x}}_\lambda$ as in Eq. (1) (There are standard package solving lasso in both matlab and python. You can also write the algorithm by yourself.). Denote

$$R = \frac{1}{d} \|\hat{\boldsymbol{x}}_\lambda - \boldsymbol{x}_0\|_2^2.$$

Make an error bar plot of $(\hat{\mathbb{E}}[R], \widehat{\mathrm{std}}[R])$ versus λ (λ will be the x-axis), with $\hat{\mathbb{E}}[R]$ the mean variable and $\widehat{\mathrm{std}}[R]$ the error variable. Please plot the curve for $d = 50, 200, 800$ on the same figure. Moreover, plot the curve of the theoretical prediction as in Theorem 1 on the same figure. The matlab code to generate the theoretical prediction is provided.

You are expected to see a U-shaped curve for this exercise.

Question 2

Fix $d = 800$, $\lambda = 1$. Generate an instance $(\boldsymbol{A}, \boldsymbol{x}_0, \boldsymbol{y})$. Compute the LASSO estimator $\hat{\boldsymbol{x}}_\lambda$. Perform a histogram plot of the non-zero coordinates of $\hat{\boldsymbol{x}}_\lambda$. Moreover, sample $(x_i)_{i \in [d]} \sim_{iid} \hat{X}$, where $\hat{X}$ is distributed as $\eta(X_0 + \tau Z; \alpha_\star \tau_\star)$, for $X_0 \sim \mathbb{P}_0$ and $Z \sim \mathcal{N}(0, 1)$ independent. Perform a histogram plot of $(x_i)_{i \in [d]}$ on the same figure.

You are expected to see that these two histogram plots are very close.

Exercise 3: Play with the LASSO risk asymptotic formula (Extra credit)

Follow the settings of Exercise 2. Define the asymptotic risk function by

$$R(\mathbb{P}_0, \lambda, \delta, \sigma^2) \equiv \mathbb{E}_{(X_0, Z) \sim \mathbb{P}_0 \times \mathcal{N}(0,1)}[(\eta(X_0 + \tau_\star Z; \theta_\star) - X_0)^2],$$

where η, $\tau_\star$, and $\alpha_\star$ are defined in Exercise 2.

Please plot the following figures:

- Take σ^2 and $\mathbb{P}_0$ to be the same as in Exercise 2. Define

  $$\lambda_\star(\delta) = \arg\min_\lambda R(\mathbb{P}_0, \lambda, \delta, \sigma^2).$$

  Plot $\lambda_\star(\delta)$ and $R(\mathbb{P}_0, \lambda_\star(\delta), \sigma^2)$ versus δ, for $\delta \in [0.1, 10]$.

- Define $\mathbb{P}_{0,\varepsilon}$ as

  $$\mathbb{P}_{0,\varepsilon}(x = 1) = \mathbb{P}_{0,\varepsilon}(x = -1) = \varepsilon, \qquad \mathbb{P}_{0,\varepsilon}(x = 0) = 1 - 2\varepsilon.$$

  Fix $\delta = 0.64$ and $\sigma^2 = 0.2$. Plot $R(\mathbb{P}_{0,\varepsilon}, \lambda_\star(\delta), \sigma^2)$ versus ε, for $\varepsilon \in [0, 0.064]$.

Theoretical exercises

Exercise 4. The exponential family calculus

In this exercise, we don't go into measure-theoretic issues. You can assume every function is integrable, and integration and derivative can be exchanged.

Let $\mu_{\mathbf{0}}$ be a reference probability measure on Ω. Let $\boldsymbol{M} : \Omega \to \mathbb{R}^d$ be any measurable function. For $\boldsymbol{\lambda} \in \mathbb{R}^d$, we denote the Gibbs-measure $\mu_{\boldsymbol{\lambda}}$ by

$$\mu_{\boldsymbol{\lambda}}(\mathrm{d}\boldsymbol{\sigma}) \propto \exp\{\langle \boldsymbol{\lambda}, \boldsymbol{M}(\boldsymbol{\sigma})\rangle\}\mu_{\mathbf{0}}(\mathrm{d}\boldsymbol{\sigma}).$$

You can think about $\mu_{\boldsymbol{\lambda}}$ as the Gibbs measure defined in class with inverse temperature $\beta = 1$. Here the Hamiltonian is absorbed in $\mu_{\mathbf{0}}$.

We further denote the log normalizing constant (the free entropy) by

$$\Phi(\boldsymbol{\lambda}) = \log \int_\Omega \exp\{\langle \boldsymbol{\lambda}, M(\boldsymbol{\sigma})\rangle\}\mu_{\mathbf{0}}(\mathrm{d}\boldsymbol{\sigma}),$$

and for any measurable function $f : \Omega \to \mathbb{R}$, we denote

$$\langle f \rangle_{\boldsymbol{\lambda}} \equiv \int_\Omega f(\boldsymbol{\sigma})\mu_{\boldsymbol{\lambda}}(\mathrm{d}\boldsymbol{\sigma}).$$

Question: Please show that

$$\nabla_{\boldsymbol{\lambda}}\Phi(\boldsymbol{\lambda}) = \langle \boldsymbol{M} \rangle_{\boldsymbol{\lambda}},$$

and

$$\nabla^2_{\boldsymbol{\lambda}}\Phi(\boldsymbol{\lambda}) = \langle \boldsymbol{M}\boldsymbol{M}^\mathsf{T} \rangle_{\boldsymbol{\lambda}} - \langle \boldsymbol{M} \rangle_{\boldsymbol{\lambda}}\langle \boldsymbol{M}^\mathsf{T} \rangle_{\boldsymbol{\lambda}} = \langle (\boldsymbol{M} - \langle \boldsymbol{M} \rangle_{\boldsymbol{\lambda}})(\boldsymbol{M} - \langle \boldsymbol{M} \rangle_{\boldsymbol{\lambda}})^\mathsf{T} \rangle_{\boldsymbol{\lambda}}.$$

Remark: In statistics, these formulas are the calculus of the exponential family. The matrix $\nabla^2_{\boldsymbol{\lambda}} Z(\boldsymbol{\lambda})$ is the Fisher-information matrix in estimating the parameter $\boldsymbol{\lambda}$. These formulas are very useful in this course.

Exercise 5: Concentration of the log-determinant of random matrices

Let $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ with $X_{ij} \sim \mathcal{N}(0, 1/d)$. Let $\lambda > 0$ and $n = \gamma d$ for fixed $\gamma \in (0, \infty)$. Denote

$$G_n(\lambda, \boldsymbol{X}) \equiv \frac{1}{n}\log(\det(\lambda \mathbf{I}_n + \boldsymbol{X}\boldsymbol{X}^\mathsf{T})).$$

Question: Please show that, for any $\lambda_0 > 0$, $\gamma \in (0, \infty)$, and $\varepsilon > 0$, we have

$$\lim_{n\to\infty, n/d\to\gamma} \mathbb{P}\Big(\sup_{\lambda\in[\lambda_0,\infty)} \Big| G_n(\lambda, \boldsymbol{X}) - \mathbb{E}[G_n(\lambda, \boldsymbol{X})] \Big| \geq \varepsilon \Big) = 0.$$

Hint: First, for fixed $\lambda > 0$, show that G_n is Lipschitz in $\boldsymbol{X}$, and use Gaussian concentration inequality to give a non-asymptotic concentration bound (remember to rescale $\boldsymbol{X}$). Then use the ε-net method to take union bound over λ in a compact interval. The non-compact part of λ can be dealt with by showing that, for any $\varepsilon > 0$, there exists $u < \infty$, such that

$$\lim_{n\to\infty} \mathbb{P}\Big(\sup_{\lambda\geq u} \Big| G_n(\lambda) - \log(\lambda) \Big| \geq \varepsilon \Big) = 0,$$

and

$$\lim_{n\to\infty} \sup_{\lambda\geq u} \Big| \mathbb{E}[G_n(\lambda)] - \log(\lambda) \Big| \leq \varepsilon.$$

Exercise 6. Concentration of the overlap of the Bayes Posterior

Let $\boldsymbol{v} \sim \nu_n = \mathrm{Unif}(\mathbb{S}^{n-1})$ be the signal vector. We observe a matrix $\boldsymbol{Y} = \lambda \boldsymbol{v}\boldsymbol{v}^\mathsf{T} + \boldsymbol{W}$, where $\boldsymbol{W} \sim \mathrm{GOE}(n)$ is independent of $\boldsymbol{v}$. The parameter λ is the signal-to-noise ratio and is a fixed real number that does not grow with n. Given observation $\boldsymbol{Y}$, we would like to estimate $\boldsymbol{v}$. Let us consider the posterior mean estimator $\widehat{\boldsymbol{V}} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$:

$$\widehat{\boldsymbol{V}}(\boldsymbol{Y}) = \mathbb{E}[\boldsymbol{v}\boldsymbol{v}^\mathsf{T}|\boldsymbol{Y}] = \int \boldsymbol{\sigma}\boldsymbol{\sigma}^\mathsf{T} p(\mathrm{d}\boldsymbol{\sigma}|\boldsymbol{Y}),$$

where $p(\mathrm{d}\boldsymbol{\sigma}|\boldsymbol{Y})$ is the posterior distribution of $\boldsymbol{v}$ given the observation matrix $\boldsymbol{Y}$ (show this by yourself):

$$p(\mathrm{d}\boldsymbol{\sigma}|\boldsymbol{Y}) \propto \exp\{n\lambda\langle\boldsymbol{\sigma}, \boldsymbol{Y}\boldsymbol{\sigma}\rangle/2\}\nu_n(\mathrm{d}\boldsymbol{\sigma}).$$

The reason why we work with $\mathbb{E}[\boldsymbol{v}|\boldsymbol{Y}]$ instead of $\mathbb{E}[\boldsymbol{v}\boldsymbol{v}^\mathsf{T}|\boldsymbol{Y}]$ is that, $\mathbb{E}[\boldsymbol{v}|\boldsymbol{Y}]$ is always 0 and is an uninformative estimator.

In this exercise, we will show the concentration of the overlap $q(\boldsymbol{Y}) \equiv \langle \boldsymbol{v}, \widehat{\boldsymbol{V}}(\boldsymbol{Y})\boldsymbol{v}\rangle$ (the larger the overlap, the better the estimator). We decompose this task to several steps.

First, motivated by the free energy trick, we define the normalized free energy as(up to a sign)

$$F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W}) = \frac{1}{n\beta}\log\int_{\mathbb{S}^{n-1}} \exp\{n\beta\langle\boldsymbol{\sigma},\boldsymbol{Y}\boldsymbol{\sigma}\rangle/2\}\nu_n(\mathrm{d}\boldsymbol{\sigma}) = \frac{1}{n\beta}\log\int_{\mathbb{S}^{n-1}} \exp\{n\beta\langle\boldsymbol{\sigma},(\lambda\boldsymbol{v}\boldsymbol{v}^\mathsf{T}+\boldsymbol{W})\boldsymbol{\sigma}\rangle/2\}\nu_n(\mathrm{d}\boldsymbol{\sigma}).$$

Question 1: Show that for any compact intervals $\mathcal{B}, \Lambda \subset (0,\infty)$, we have (the probability and expectation is taken with respect to $\boldsymbol{v}$ and $\boldsymbol{W}$)

$$\lim_{n\to\infty} \mathbb{P}\Big(\sup_{\lambda\in\Lambda,\beta\in\mathcal{B}} \Big|F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W}) - \mathbb{E}[F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})]\Big| \geq \varepsilon\Big) = 0.$$

Hint: Follow the similar steps as in Exercise 5. Note that by symmetry, the probability and the expectation does not depend on the choice of $\boldsymbol{v}$.

Question 2: Show that for any $\lambda \in (0,\infty)$, we have

$$\partial_\lambda F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})|_{\beta=\lambda} = \langle\boldsymbol{v}, \widehat{\boldsymbol{V}}(\boldsymbol{Y})\boldsymbol{v}\rangle/2,$$

and $F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})$ is convex in λ:

$$\partial_\lambda^2 F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W}) \geq 0.$$

Hint: Use the calculus of the exponential family.

Question 3: Show the following. Suppose

$$f(\beta,\lambda) \equiv \lim_{n\to\infty} \mathbb{E}[F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})]$$

exists. Moreover, suppose there exists a compact interval $\Lambda \subset (0,\infty)$, such that for any $\beta \in \Lambda$, $\partial_\lambda f(\beta,\lambda)$ is continuous with respect to $\lambda \in \Lambda$. Then for any $\varepsilon > 0$, we have (the probability and expectation is with respect to $\boldsymbol{v}$ and $\boldsymbol{W}$)

$$\lim_{n\to\infty} \mathbb{P}\Big(\Big|\langle\boldsymbol{v}, \widehat{\boldsymbol{V}}(\boldsymbol{Y})\boldsymbol{v}\rangle/2 - \partial_\lambda f(\beta,\lambda)|_{\beta=\lambda}\Big| \geq \varepsilon\Big) = 0.$$

Hint: Prove a quantitative version of the following lemma, and take $f_n(\lambda) = F_n(\beta,\lambda,\boldsymbol{v},\boldsymbol{W})$ for fixed β, $\boldsymbol{v}$ and $\boldsymbol{W}$.

Lemma 1. *Let $\{f_n(\lambda)\}_{n\geq 0}$ be a deterministic sequence of* **convex** *functions on a $[0,1]$. Suppose the function converges to some convex function f pointwisely:*

$$\lim_{n\to\infty} |f_n(\lambda) - f(\lambda)| = 0.$$

and suppose $f'(\lambda)$ exists for any $\lambda \in (0,1)$. Then its derivative also converges to f' pointwisely on $\lambda \in (0,1)$:

$$\lim_{n\to\infty} |f_n'(\lambda) - f'(\lambda)| = 0.$$

Remark: The whole proof can be generalized to the case when ν_n is any probability measure that is supported on the unit ball $\mathsf{B}(0,1)$, and the Bayes estimator with mis-specified prior $\beta \neq \lambda$.

Exercise 7: The free energy trick and replicas

Let $H : \Omega_d = \{\pm 1\}^d \to \mathbb{R}$ be a Hamiltonian, and let μ_β be the associated Gibbs probability measure of H, i.e.,

$$\mu_\beta(\boldsymbol{\sigma}) \propto \exp\{-\beta H(\boldsymbol{\sigma})\}.$$

For any $\boldsymbol{g} \in \Omega_d \to \mathbb{R}^p$, define its ensemble average $\langle \boldsymbol{g} \rangle_\beta$ by

$$\langle \boldsymbol{g}(\boldsymbol{\sigma}) \rangle_\beta \equiv \sum_{\boldsymbol{\sigma} \in \Omega_d} \boldsymbol{g}(\boldsymbol{\sigma}) \mu_\beta(\boldsymbol{\sigma}).$$

Question 1

Find a Hamiltonian $H_{\boldsymbol{\lambda}}$ on the space Ω_d, parameterized by $\boldsymbol{\lambda} \in \mathbb{R}^d$, such that defining

$$F(\beta, \boldsymbol{\lambda}) = -\frac{1}{\beta} \log \sum_{\boldsymbol{\sigma} \in \Omega_d} \exp\{-\beta H_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})\},$$

we have $\nabla_{\boldsymbol{\lambda}} F(\beta, \boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\mathbf{0}} = \langle \boldsymbol{\sigma} \rangle_\beta$.

Question 2

Find a Hamiltonian H_λ on some space Ω_d, parameterized by $\lambda \in \mathbb{R}$, such that defining

$$F(\beta, \lambda) = -\frac{1}{\beta} \log \sum_{\boldsymbol{\sigma} \in \Omega_d} \exp\{-\beta H_\lambda(\boldsymbol{\sigma})\},$$

we have $\nabla_\lambda F(\beta, \lambda)|_{\lambda=0} = \langle \|\boldsymbol{\sigma}\|_2^2 \rangle_\beta$.

Question 3 (Extra credit)

Find a Hamiltonian H_λ on some space Ω', parameterized by $\lambda \in \mathbb{R}$, such that defining

$$F(\beta, \lambda) = -\frac{1}{\beta} \log \sum_{\boldsymbol{\sigma}' \in \Omega'} \exp\{-\beta H_\lambda(\boldsymbol{\sigma}')\},$$

we have $\nabla_\lambda F(\beta, \lambda)|_{\lambda=0} = \|\langle \boldsymbol{\sigma} \rangle_\beta\|_2^2$. (Hint: This is different from $\langle \|\boldsymbol{\sigma}\|_2^2 \rangle_\beta$. Consider to take $\Omega' = \Omega_d \times \Omega_d$.)

Question 4 (Extra credit)

Let $\boldsymbol{A} \in (\mathbb{R}^d)^{\otimes p}$ be a p-tensor. Please find a Hamiltonian H_λ on some space Ω', parameterized by $\lambda \in \mathbb{R}$, such that defining

$$F(\beta, \lambda) = -\frac{1}{\beta} \log \sum_{\boldsymbol{\sigma}' \in \Omega'} \exp\{-\beta H_\lambda(\boldsymbol{\sigma}')\},$$

we have $\nabla_\lambda F(\beta, \lambda)|_{\lambda=0} = \langle \boldsymbol{A}, \langle \boldsymbol{\sigma} \rangle_\beta^{\otimes p} \rangle$ (the outer bracket $\langle \cdot, \cdot \rangle$ is the inner product in tensor space, while the inner bracket $\langle \cdot \rangle_\beta$ is the ensemble average under measure μ_β).