Final Examination

Exercise 1 (theory):

We consider a data set made of M real-valued and $L-$dimensional vectors, $\mathbf{s}^m = (s_1^m, s_2^m, ..., s_L^m)$, with $m = 1...M$. We define the first and the second moments of the variables computed over the data

$$\langle s_i \rangle^d = \frac{1}{M} \sum_{m=1}^{M} s_i^m \quad , \qquad \langle s_i s_j \rangle^d = \frac{1}{M} \sum_{m=1}^{M} s_i^m s_j^m \ . \tag{1}$$

Hereafter, we assume in **ALL** questions **EXCEPT 1C** that the first moments vanish: $\langle s_i \rangle^d = 0$. The objective of this exercise is to infer a Gaussian multivariate density of probability p from those data, then to characterize the distribution of the log-likelihoods $\log p(\mathbf{s})$ over all possible configurations $\mathbf{s}$. Question 3 can be done independently of the previous questions.

Question 1. Write down the general expression of a multivariate Gaussian density of probability $p(\mathbf{s})$ over variables having zero means and covariances $C_{ij} = \langle s_i \, s_j \rangle - \langle s_i \rangle \langle s_j \rangle$.

A. Recall how to infer the values of the C_{ij}'s that maximises the log-likelihood of the data (MLE).

B. What is the class of conjugate prior to the likelihood p? How would the previous result for C_{ij} be modified if you used maximum a posteriori inference (MAP) with such a conjugate prior instead of MLE?

C. How would you infer p within MLE if the averages values $\langle s_i \rangle^d$ were not equal to zero?

Question 2. We introduce the distribution ρ of the log-likelihoods ℓ of configurations drawn from p inferred with MLE,

$$\rho(\ell) = \int d\mathbf{s} \; p(\mathbf{s}) \; \delta\big(\ell - \log p(\mathbf{s})\big) \ . \tag{2}$$

The moment of order k of the log-likelihood is defined as

$$\langle \ell^k \rangle = \int d\ell \, \rho(\ell) \, \ell^k \ . \tag{3}$$

Our goal is to determine the smallest value of the log-likelihood for a set of configurations sampled from p as a function of the size of this set.

A. Compute the average log-likelihood $\langle \ell \rangle$.

B. Show that the variance of the log-likelihood is equal to

$$\mathrm{var}(\ell) \equiv \langle \ell^2 \rangle - \langle \ell \rangle^2 = \frac{L}{2} \ . \tag{4}$$

Hint: consider first the case $L = 1$. We recall that

$$\int dx \, \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, x^2 = 1 \quad , \qquad \int dx \, \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, x^4 = 3 \ . \tag{5}$$

To extend the result to $L \geq 2$ move to the eigenbasis of C for integrating over the L–dimensional configurations $\mathbf{s}$.

Question 3. We now approximate the distribution of the log-likelihoods by the following Gaussian density of probability:

$$\rho(\ell) \simeq \frac{1}{\sqrt{2\pi\,\mathrm{var}(\ell)}} \exp\left(-\frac{(\ell - \langle \ell \rangle)^2}{2\,\mathrm{var}(\ell)}\right) \ . \tag{6}$$

A. Let us draw randomly and independently M samples $\ell_1, \ell_2, ..., \ell_M$ from the distribution in eqn. (6), and call $\ell_{\min}$ their minimum. Show that the probability $P(\lambda)$ that $\ell_{min} > \lambda$ is given by

$$P(\lambda) = \left(\int_\lambda^\infty d\ell\, \rho(\ell)\right)^M \ . \tag{7}$$

Sketch the representative curve of $P(\lambda)$ as a function of λ when M is large. Briefly justify why one can expect that $\langle \ell \rangle - \ell_{min}$ is likely to be much larger than $\sqrt{\mathrm{var}(\ell)}$ when M is large.

B. Deduce that we can approximate

$$P\left(\lambda = \langle \ell \rangle - u\sqrt{\mathrm{var}(\ell)}\right) = \left(1 - \int_u^\infty dx\, \frac{e^{-x^2/2}}{\sqrt{2\pi}}\right)^M \simeq \exp\left(-M \int_u^\infty dx\, \frac{e^{-x^2/2}}{\sqrt{2\pi}}\right) \ . \tag{8}$$

Using the asymptotic expression for the complementary error function for large positive u,

$$\int_u^\infty dx\, \frac{e^{-x^2/2}}{\sqrt{2\pi}} \simeq \frac{e^{-u^2/2}}{u\,\sqrt{2\pi}} \ , \tag{9}$$

check that the expression for P in eqn. (8) is qualitatively compatible with the sketch in the previous question.

C. Conclude that, with high probability when M is large and to dominant order in M,

$$\ell_{min} \simeq \langle \ell \rangle - \sqrt{2 \log M\, \mathrm{var}(\ell)} \ . \tag{10}$$

Question 4. Discuss the validity of the Gaussian approximation in eqn. (6).