Exercises on Chapter 3: Statistics and concentration inequalities

Mathematics of Data Science, M1 IDD

November-December 2024*

Dauphine | PSL 🔀

Exercise 3.1: With and without concentration inequalities

Suppose that we toss a fair coin (i.e. that has probability $\frac{1}{2}$ of landing on heads or tails) N times in an independent fashion. Let h_N be the number of times we obtain heads.

- a) Shown that $\mathbb{E}[h_N] = \frac{N}{2}$ and $\operatorname{Var}[h_N] = \frac{N}{4}$. Hint: Use the fact that if x and y are two independent random variables, then $\mathbb{E}[x+y] = \mathbb{E}[x] + \mathbb{E}[y]$ and $\operatorname{Var}[x+y] = \operatorname{Var}[x] + \operatorname{Var}[y]$.
- b) Apply Chebyshev's inequality to bound the probability of getting at least $\frac{3N}{4}$ heads.
- c) For this particular problem, one can derive the following Hoeffding-type inequality¹:

$$\mathbb{P}(h_N \ge t) \le \exp\left[-\frac{(2t-N)^2}{2N}\right].$$

Using this inequality, provide another bound on the probability of getting at least $\frac{3N}{4}$ heads. Compare this inequality with that of question b).

^{*}Version 4, last updated December 10, 2024.

¹To be described in class.

Exercise 3.2: Chernoff inequalities

In this exercise, we study another type of concentration inequalities than that seen in class called *Chernoff bounds* or *Chernoff inequalities*. In the general form, this inequality states that for any random variable y and any $t \in \mathbb{R}$, we have

$$\mathbb{P}(y \ge t) \le \min_{\lambda \ge 0} \mathbb{E}\left[\exp(\lambda(y-t))\right].$$
(1)

a) Proving (1) amounts to proving

$$\ln\left(\mathbb{P}\left(y \ge t\right)\right) \le \min_{\lambda \ge 0} \ln\left(\mathbb{E}\left[\exp(\lambda(y-t))\right]\right).$$
(2)

Justify that right-hand side of (2) is the solution to a convex optimization problem. To this end, you may use a generalization of the Hölder inequality from Exercise 1.8, that states that for any random variables w, z, we have

$$\mathbb{E}_{w,z}[w\,z] \le \mathbb{E}_w[|w|^p]^{1/p} \mathbb{E}_z[|z|^q]^{1/q}$$

any pair (p,q) such that p>1,q>1 and $\frac{1}{p}+\frac{1}{q}=1.$

b) Suppose that $y \sim \mathcal{N}(0,1)$. In that case, one can show that $\ln (\mathbb{E}[\exp(\lambda y)]) = \frac{\lambda^2}{2}$. Use this property to deduce from (1) that

$$\mathbb{P}(y \ge t) \le \exp\left(-\frac{t^2}{2}\right)$$

for any t > 0. What inequality do you obtain for $t \le 0$?

Exercise 3.3: Boosting

Suppose that we perform 2m independent runs of a randomized algorithm designed to solve a decision problem (e.g. is a given convex optimization problem feasible?). Because of the randomness, the algorithm is only correct with probability $\frac{1}{2} + \delta$ for some $\delta \in (0, 1)$. To make a decision, we choose the output returned by the majority of runs.

- a) Let y_i be a Bernoulli random variable such that $y_i = 1$ if the *i*th run returns the wrong output, and $y_i = 0$ otherwise. Compute $\mathbb{E}[y_i]$.
- b) Express the probability of making the right conclusion from the output of the 2m instances.
- c) Let $p \in [0,1)$. Using Hoeffding's inequality, show that the probability of making the right conclusion is at least 1-p when

$$m \ge \frac{1}{4\delta^2} \ln\left(\frac{1}{p}\right).$$

Exercise 3.4: Chernoff inequalities for vectors

In this exercise, we seek a Chernoff-type bound in a vector setting. More precisely, we consider a Gaussian vector $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}_{\mathbb{R}^n}, \boldsymbol{I}_n)$ and a nonempty polyhedral set defined by $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ with $\boldsymbol{A} \in \mathbb{R}^{\ell \times n}$ and $\boldsymbol{b} \in \mathbb{R}^{\ell}$. Our goal is to provide a bound of the form

$$\mathbb{P}\left(\boldsymbol{y}\in\mathcal{C}\right)\leq\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}+\boldsymbol{\mu}\right)\right]$$
(3)

where $\lambda \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$. As in the previous exercise, we would like to obtain the tightest bound possible.

- a) Using that $\mathbb{P}(\boldsymbol{y} \in \mathcal{C}) = \mathbb{E}[1_{\mathcal{C}}(\boldsymbol{y})]$, justify that any pair $(\boldsymbol{\lambda}, \mu) \in \mathbb{R}^n \times \mathbb{R}$ satisfying $\exp(\boldsymbol{\lambda}^T \boldsymbol{y} + \mu) \ge 1_{\mathcal{C}}(\boldsymbol{y})$ for every $\boldsymbol{y} \in \mathbb{R}^n$ also satisfies (3) with $-\boldsymbol{\lambda}^T \boldsymbol{y} \le \mu \ \forall \boldsymbol{y} \in \mathcal{C}$.
- b) By considering logarithms, show that

$$\ln\left(\mathbb{P}\left(\boldsymbol{y}\in\mathcal{C}\right)\right)\leq\min_{\boldsymbol{\lambda}\in\mathbb{R}^{n}}\left\{S_{\mathcal{C}}(-\boldsymbol{\lambda})+\ln\mathbb{E}\left[e^{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{z}}\right]\right\},$$
(4)

with $S_{\mathcal{C}}: \boldsymbol{y} \mapsto \max_{\boldsymbol{x} \in \mathcal{C}} \boldsymbol{y}^{\mathrm{T}} \boldsymbol{x}$.

c) Since \boldsymbol{y} is Gaussian, we have that $\ln(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]) = \frac{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda}}{2}$ for any $\boldsymbol{\lambda}$. In addition, we can show that

$$S_{\mathcal{C}}(oldsymbol{y}) = \min_{oldsymbol{u} \in \mathbb{R}^\ell} \left\{ oldsymbol{b}^{\mathrm{T}} oldsymbol{u} ig| oldsymbol{A}^{\mathrm{T}} oldsymbol{u} = oldsymbol{y}, oldsymbol{u} \geq oldsymbol{0}
ight\}$$

for any $y \in \mathbb{R}^n$. Show then that the right-hand side of (4) corresponds to the optimal value of the quadratic problem

minimize<sub>$$\boldsymbol{\lambda} \in \mathbb{R}^{n}, \boldsymbol{v} \in \mathbb{R}^{\ell}$$
 $\boldsymbol{b}^{\mathrm{T}} \boldsymbol{v} + \frac{\|\boldsymbol{\lambda}\|^{2}}{2}$
s.t. $\boldsymbol{v} \geq \mathbf{0},$ (5)
 $\boldsymbol{A}^{\mathrm{T}} \boldsymbol{v} + \boldsymbol{\lambda} = \mathbf{0}.$</sub>

d) The problem (5) is equivalent to

$$\min_{\boldsymbol{v}\in\mathbb{R}^{\ell}}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{v} + \frac{\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v}\|^{2}}{2} \quad \text{s.t.} \quad \boldsymbol{v}\geq\boldsymbol{0}, \tag{6}$$

where we reformulated the problem so as to eliminate the λ variables while preserving the same optimal value.

i) Using that same reformulation technique, show that the dual of problem (6) is equivalent to

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{x} \in \mathbb{R}^m} & -\frac{\|\boldsymbol{x}\|^2}{2} \\ \text{s.t.} & \boldsymbol{A} \boldsymbol{x} \leq \boldsymbol{b}. \end{array}$$
(7)

- ii) Justify that the optimal value of problem (7) is $-\frac{1}{2}\text{dist}(\mathbf{0}, C)^2$, where $\text{dist}(\mathbf{a}, C) = \min_{\mathbf{y} \in C} \|\mathbf{y} \mathbf{a}\|$.
- iii) Strong duality holds for problem (6). Using this property, provide a closed-form expression for (4) and (3).

Exercise 3.5: Random projections (Exam 2023-2024)

In this exercise, we consider projections on random subspaces. Letting $r \leq n$, we define a random projection matrix $\mathbf{P} \in \mathbb{R}^{r \times n}$ such that the coefficients of \mathbf{P} are i.i.d. and follow a normal distribution $\mathcal{N}(0, \frac{1}{r})$. Recall that the probability density function associated with $y \sim \mathcal{N}(0, \frac{1}{r})$ is

$$p(y) = \sqrt{\frac{r}{2\pi}} \exp\left(-\frac{ry^2}{2}\right) \qquad \forall y \in \mathbb{R}$$

- a) Give a formula for the probability density function of a column of P (e.g. the first one).
- b) Show that this density is log-concave.
- c) Given any vector $a \in \mathbb{R}^n$ and any tolerance $\epsilon \in (0,1]$, it can be shown (using Johnson-Lindenstrauss-type arguments) that

$$\mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{a}\| \le (1-\epsilon)\|\boldsymbol{a}\|\right) \le \exp\left(-c\,r\epsilon^2\right),\tag{8}$$

where c > 0 is a constant independent of r, n and ϵ .

i) Using (8), find a sufficient condition for the bound

$$\mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{a}\| > (1-\epsilon)\|\boldsymbol{a}\|\right) \ge 0.99\tag{9}$$

to hold.

- ii) Use the condition from the previous question to determine \bar{r} such that (9) holds whenever $r \geq \bar{r}$.
- iii) Suppose that we cannot generate matrices using $r = \bar{r}$, but that we can generate matrices $P \in \mathbb{R}^{\underline{r} \times n}$ with $\underline{r} < \bar{r}$ such that

$$\mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{a}\| > (1-\epsilon)\|\boldsymbol{a}\|\right) \ge 0.6.$$

How many such matrices would be necessary to ensure that one of them satisfies $\|\mathbf{P}\mathbf{a}\| > (1-\epsilon)\|\mathbf{a}\|$ with probability at least 0.99?

Exercise 3.6: Erdös-Rényi graphs

Graphs generated using the Erdös-Rényi model G(n, p) are undirected random graphs with n vertices. For every *possible* edge (i, j), the probability that (i, j) is an edge of the graph is p, independently of the other edges.

- a) For any vertex $i \in \{1, ..., n\}$, we let d_i denote the degree of this vertex, that is the number of edges that include that vertex. Express d_i as a sum of n-1 independent Bernoulli variables.
- b) Show then that $\mathbb{E}[d_i] = \overline{d} := (n-1)p$.

c) A version of Chernoff's inequality states that if x_1, \ldots, x_N are independent Bernoulli variables with parameters p_1, \ldots, p_N , then

$$\forall \delta \in (0,1], \qquad \mathbb{P}\left(\left|\sum_{j=1}^{N} x_j - \mu\right| \ge \delta \mu\right) \le 2\exp(-c\mu\delta^2), \tag{10}$$

where $\mu = \mathbb{E}\left[\sum_{j=1}^{N} x_j\right]$ and c > 0. Use (10) to show that for any $i = 1, \ldots, n$, we have

$$\mathbb{P}\left(|d_i - \bar{d}| \ge 0.1d\right) \le 2\exp(-C\bar{d}) \tag{11}$$

where C > 0 is a universal constant.

d) Use the inequality (11) to find a bound on

$$\mathbb{P}\left(\exists i \in \{1, \dots, n\}, \quad |d_i - \bar{d}| \ge 0.1\bar{d}\right).$$

e) Conclude that there exists a constant $\hat{c} > 0$ such that

$$\mathbb{P}\left(\max_{1 \leq i \leq n} |d_i - \bar{d}| < 0.1d\right) \geq 0.9 \quad \text{when} \quad p \geq \hat{c} \frac{\ln(n)}{n-1}.$$

NB: This result shows that when $\bar{d} = O(\ln(n))$, the degrees of all vertices is approximately equal to \bar{d} .

Solutions

Solution for Exercise 3.1: Without concentration inequalities

a) Let x_n be a random variable equal to 1 if we got heads on the *n*th throw and 0 otherwise. Then $h_N = \sum_{i=1}^N x_n$, and by independence, we have:

$$\mathbb{E}[h_N] = \sum_{n=1}^N \mathbb{E}[x_n]$$

=
$$\sum_{n=1}^N (\mathbb{P}(x_n = 1) \times 1 + \mathbb{P}(x_n = 0) \times 0)$$

=
$$\sum_{n=1}^N \mathbb{P}(x_n = 1) = \frac{N}{2}.$$

Similarly, noting that $\mathbb{E}[x_n] = \frac{1}{2}$, we

$$\operatorname{Var}[h_N] = \sum_{n=1}^{N} \operatorname{Var}[x_n]$$
$$= \sum_{n=1}^{N} (\mathbb{E}[x_n^2] - \mathbb{E}[x_n]^2)$$
$$= \sum_{n=1}^{N} \left(\mathbb{P}(x_n = 1) - \frac{1}{4} \right)$$
$$= \sum_{n=1}^{N} \frac{1}{4} = \frac{N}{4}.$$

b) Chebyshev's inequality applied to a random variable x states that

$$\mathbb{P}\left(\left|x-\mathbb{E}\left[x\right]\right| \geq t\right) \leq \frac{\mathrm{Var}\left[x\right]}{t^{2}}$$

for any t > 0. When applied to h_N , this inequality becomes

$$\mathbb{P}\left(|h_N - \frac{N}{2}| \ge t\right) \le \frac{N}{4t^2}$$

Our probability of interest is $\mathbb{P}\left(h_N \geq \frac{3N}{4}\right)$, that satisfies

$$\mathbb{P}\left(h_N \ge \frac{3N}{4}\right) = \mathbb{P}\left(h_N - \frac{N}{2} \ge \frac{N}{4}\right) \le \mathbb{P}\left(|h_N - \frac{N}{2}| \ge \frac{N}{4}\right).$$

Using Chebyshev's inequality with $t = \frac{N}{4}$, we arrive at

$$\mathbb{P}\left(h_N \ge \frac{3N}{4}\right) \le \frac{4}{N}.$$

c) Using $t = \frac{3N}{4}$ immediately gives

$$\mathbb{P}\left(h_N \ge \frac{3N}{4}\right) \le \exp\left[-\frac{N}{8}\right].$$

As a result, the bound obtained through Hoeffding's inequality goes significantly faster to 0 than the bound obtained through Chebyshev's inequality (exponentially fast rather than linearly fast).

Solution for Exercise 3.2: Chernoff inequality

Foreword to question a) The equivalence between (1) and (2) can be justified as follows. Suppose that $\lambda \ge 0$ satisfies $\mathbb{P}(y \ge t) \le \mathbb{E}[\exp(\lambda(y-t))]$ for all $t \in \mathbb{R}$. Then, by taking logarithms on both sides of the inequality (allowing $\ln(0) = -\infty$ and $-\infty \le -\infty$), we obtain

$$\ln\left(\mathbb{P}\left(y \ge t\right)\right) \le \ln\left(\mathbb{E}\left[\exp\left(\lambda(y-t)\right)\right]\right).$$

It remains to show that minimizing the right-hand side of the latter inequality over λ gives the same bound that that obtained by (1). For any $\mu \geq 0$, using that the exponential function is monotonically increasing gives

$$\exp\left[\min_{\lambda\geq 0}\ln\left(\mathbb{E}\left[\exp\left(\lambda(y-t)\right)\right]\right)\right] \leq \exp\left[\ln\left(\mathbb{E}\left[\exp\left(\mu(y-t)\right)\right]\right)\right] = \mathbb{E}\left[\exp\left(\mu(y-t)\right)\right].$$

Hence

$$\exp\left[\min_{\lambda \ge 0} \ln\left(\mathbb{E}\left[\exp\left(\lambda(y-t)\right)\right]\right)\right] \le \min_{\mu \ge 0} \mathbb{E}\left[\exp\left(\mu(y-t)\right)\right]$$
$$\Rightarrow \qquad \min_{\lambda \ge 0} \ln\left(\mathbb{E}\left[\exp\left(\lambda(y-t)\right)\right]\right) \le \ln\left[\min_{\mu \ge 0} \mathbb{E}\left[\exp\left(\mu(y-t)\right)\right]\right].$$

Conversely, using that $\ln(\cdot)$ is monotonically increasing gives

$$\ln [\min_{\mu \ge 0} \mathbb{E} [\exp (\mu(y-t))]] \le \ln [\mathbb{E} [\exp (\lambda(y-t))]]$$

$$\Rightarrow \ln [\min_{\mu \ge 0} \mathbb{E} [\exp (\mu(y-t))]] \le \min_{\lambda \ge 0} \ln [\mathbb{E} [\exp (\lambda(y-t))]]$$

Overall, we have shown that

$$\ln\left[\min_{\lambda\geq 0}\mathbb{E}\left[\exp\left(\mu(y-t)\right)\right]\right] = \min_{\lambda\geq 0}\ln\left[\mathbb{E}\left[\exp\left(\lambda(y-t)\right)\right]\right],$$

and therefore the two equalities are equivalent.

Question a) Observe that

$$\min_{\lambda \ge 0} \ln \left(\mathbb{E} \left[\exp \left(\lambda (y - t) \right) \right] \right) = \min_{\lambda \ge 0} \ln \left(\mathbb{E} \left[\exp(\lambda y) \right] \right) - \lambda t.$$

Since $t \mapsto -\lambda t$ is a linear function of λ , it is convex, and thus it suffices to show that $\lambda \mapsto \ln (\mathbb{E} [\exp(\lambda y)])$ is convex on \mathbb{R}_+ to arrive at the desired result.

To this end, we consider two values $\lambda_1 \ge 0$ and $\lambda_2 \ge 0$, as well as $\alpha \in [0,1]$. Our goal is to prove

$$\ln\left(\mathbb{E}\left[\exp((\alpha\lambda_1 + (1-\alpha)\lambda_2)y)\right]\right) \le \alpha \ln\left(\mathbb{E}\left[\exp(\lambda_1 y)\right]\right) + (1-\alpha)\ln\left(\mathbb{E}\left[\exp(\lambda_2 y)\right]\right)$$

If $\alpha \in \{0, 1\}$, the result trivially holds. Otherwise, we apply Minkowski's inequality to $\mathbb{E} \left[\exp((\alpha \lambda_1 + (1 - \alpha)\lambda_2) \mathbb{E} \left[\exp(\alpha \lambda_1 y) \times \exp((1 - \alpha)\lambda_2 y) \right] \right]$. Using $Y = \alpha \lambda_1 y$, $Z = (1 - \alpha)\lambda_2 y$, $p = \frac{1}{\alpha}$ and $q = \frac{1}{1 - \alpha}$. We obtain

$$\mathbb{E}\left[\exp((\alpha\lambda_1 + (1-\alpha)\lambda_2)y)\right] \leq \mathbb{E}\left[\exp(\alpha\lambda_1 y)^{1/\alpha}\right]^{\alpha} \mathbb{E}\left[\exp((1-\alpha)\lambda_2 y)^{\alpha}\right]^{1-\alpha} \\ = \mathbb{E}\left[\exp(\lambda_1 y)\right]^{\alpha} \mathbb{E}\left[\exp(\lambda_2 y)\right]^{1-\alpha},$$

Taking logarithms then leads to

$$\ln\left(\mathbb{E}\left[\exp((\alpha\lambda_1 + (1-\alpha)\lambda_2)y)\right]\right) \le \alpha \ln\left(\mathbb{E}\left[\exp(\lambda_1 y)\right]\right) + (1-\alpha)\ln\left(\mathbb{E}\left[\exp(\lambda_2 y)\right]\right),$$

showing that the function is indeed convex.

Question b) By applying $\ln \mathbb{E} \left[\exp(\lambda y) \right] = \frac{\lambda^2}{2}$ in the inequality derived in question a), we obtain that

$$\ln\left[\mathbb{P}\left(y \ge t\right)\right] \le \min_{\lambda \ge 0} \left\{-\lambda t + \frac{\lambda^2}{2}\right\}$$

The objective function of the right-hand side optimization problem is a convex quadratic in λ , and its minimum is attained at $\lambda^* = \max\{t, 0\}$. Indeed, if $t \ge 0$, then the minimum is $\lambda^* = t \ge 0$, while if t < 0, we have $-\lambda t + \frac{\lambda^2}{2} \ge 0$, hence $\lambda^* \ge 0$ is a minimum². When $t \ge 0$, the inequality gives

$$\ln\left[\mathbb{P}\left(y\geq t\right)\right]]\leq-\frac{t^{2}}{2},$$

and thus $\mathbb{P}(Y \ge t) \le \exp(-t^2/2)$.

When t < 0, the inequality gives

$$\ln\left[\mathbb{P}\left(y \ge t\right)\right] \le 0,$$

hence the result becomes $\mathbb{P}(y \ge t) \le 1$, which is a correct bound (yet of little use since a probability is always bounded above by 1).

Solution for Exercise 3.3: Boosting

Question a) A straightforward calculation gives

$$\mathbb{E}[Y_i] = 1 \times \mathbb{P}(Y_i = 1) + 0 \times \mathbb{P}(Y_i = 0) = 1 - \left(\frac{1}{2} + \delta\right) = \frac{1}{2} - \delta.$$

²One can also establish this using the optimality conditions from Chapter 2.

Question b) Since the decision is made on 2m runs of the algorithm by majority voting, we know that the correct output is accepted if $\sum_{i=1}^{2m} y_i < m$, since this is only possible when more than half of the voters returned 0. As a result, the probability of making the right decision is

$$\mathbb{P}\left(\sum_{i=1}^{2m} y_i < m\right).$$

Question c) We use the variant of Hoeffding's inequality tailored to bounded random variables. ³ For any $t \ge 0$, we have

$$\mathbb{P}\left(\sum_{i=1}^{2m}\left(y_i - \mathbb{E}\left[y_i\right]\right) \ge t\right) \le \exp\left[-\frac{2t^2}{2m}\right] = \exp\left[-\frac{t^2}{m}\right],$$

where the 2m factor on the right-hand side corresponds to the squared norm of the vector of all ones. Using the formula for the expected value, we obtain

$$\mathbb{P}\left(\sum_{i=1}^{2m} y_i \ge t + m - 2m\delta\right) \le \exp\left[-\frac{t^2}{m}\right].$$

Setting $t = 2m\delta > 0$ gives

$$\mathbb{P}\left(\sum_{i=1}^{2m} y_i \ge m\right) \le \exp\left[-\frac{4m^2\delta^3}{m}\right] = \exp\left[-4m\delta^2\right]$$

Our goal is to guarantee that $\mathbb{P}\left(\sum_{i=1}^{2m} y_i < m\right) \ge 1-p$, which is equivalent to $\mathbb{P}\left(\sum_{i=1}^{2m} y_i \ge m\right) < p$. Choosing $m > \frac{1}{2\delta^2} \ln\left(\frac{1}{p}\right)$, we see that

$$\exp\left[-2m\delta^2\right] < \exp\left[-\ln\left(\frac{1}{p}\right)\right] = p,$$

and the desired conclusion follows.

Solution for Exercise 3.4: Chernoff inequalities for vectors

Question a) Consider the function

$$\begin{array}{rccc} f: & \mathbb{R}^n & \to & \mathbb{R} \\ & \boldsymbol{y} & \mapsto & \exp(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu). \end{array}$$

$$\mathbb{P}\left(\sum_{i=1}^{N} \left(y_i - \mathbb{E}\left[y_i\right]\right) \ge t\right) \le \exp\left[-\frac{2t^2}{N(M-m)^2}\right].$$

³In its general form, this inequality states that for any set of variables y_1, \ldots, y_N that are bounded in [m, M] and any $t \ge 0$, we have

such that $f(y) \ge 1_{\mathcal{C}}(y)$ for every $y \in \mathbb{R}^n$. By definition of the indicator function, it implies that

$$\begin{split} f(\boldsymbol{y}) \geq 1_{\mathcal{C}}(\boldsymbol{y}) \; \forall \boldsymbol{y} \in \mathbb{R}^{n} \; \Leftrightarrow \; \left\{ \begin{array}{ll} f(\boldsymbol{y}) \geq 1 & \forall \boldsymbol{y} \in \mathcal{C} \\ f(\boldsymbol{y}) \geq 0 & \forall \boldsymbol{y} \notin \mathcal{C} \end{array} \right. \\ \Leftrightarrow \; \left\{ \begin{array}{ll} \exp(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu) \geq 1 & \forall \boldsymbol{y} \in \mathcal{C} \\ \exp(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu)) \geq 0 & \forall \boldsymbol{y} \notin \mathcal{C} \end{array} \right. \\ \Leftrightarrow \; -\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} \leq \mu \quad \forall \boldsymbol{y} \in \mathcal{C}, \end{split}$$

where the latter equivalence comes from the fact that an exponential is always positive, hence the inequalities for $y \notin C$ always hold.

Question b) Taking logarithms on both sides of (3) gives

$$\ln\left(\mathbb{P}\left(\boldsymbol{z}\in\mathcal{C}\right)\right)\leq\ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}+\boldsymbol{\mu}\right)\right]\right).$$

We want to compute the pair (λ, μ) that yields the tightest bound. From question a), we know that such a pair must satisfy $-\lambda^T y \leq \mu \ \forall y \in C$. As a result, the best lower bound is given as an optimal value of an optimization problem over λ and μ , namely

$$\min_{\boldsymbol{\lambda},\mu} \left\{ \ln \left(\mathbb{E} \left[\exp \left(\boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} + \mu \right) \right] \right) \middle| - \boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{y} \leq \mu \,\, \forall \boldsymbol{y} \in \mathcal{C} \right\}.$$

Using $\ln \left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}+\boldsymbol{\mu}\right)\right]\right) = \ln \left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]\right) + \boldsymbol{\mu}$, the problem can be rewritten as $\min_{\boldsymbol{\lambda},\boldsymbol{\mu}} \left\{\boldsymbol{\mu}+\ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]\right) \middle| -\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} \leq \boldsymbol{\mu} \,\,\forall \boldsymbol{y} \in \mathcal{C}\right\}.$

Now, since the objective function minimizes μ and $\mu \ge -\lambda^T y \ \forall y \in C$, the optimal μ for a given λ is $\max_{y \in C} -\lambda^T y$. As a result, we can reformulate the problem as a problem involving only λ :

$$\begin{split} \min_{\boldsymbol{\lambda},\mu} \left\{ \mu + \ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]\right) \middle| - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} \leq \mu \; \forall \boldsymbol{y} \in \mathcal{C} \right\} \\ = & \min_{\boldsymbol{\lambda}} \min_{\mu} \left\{ \mu + \ln\left(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]\right) \middle| - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} \leq \mu \; \forall \boldsymbol{y} \in \mathcal{C} \right\} \\ = & \min_{\boldsymbol{\lambda}} \left\{ \max_{\boldsymbol{x} \in \mathcal{C}} [-\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{x}] + \ln(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]) \right\} \\ = & \min\left\{ S_{\mathcal{C}}(-\boldsymbol{\lambda}) + \ln(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right]) \right\}. \end{split}$$

As a result, we must have

$$\ln\left(\mathbb{P}\left(\boldsymbol{y}\in\mathcal{C}\right)\right)\leq\min_{\boldsymbol{\lambda}}\left\{S_{\mathcal{C}}(-\boldsymbol{\lambda})+\ln(\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}\right)\right])\right\}.$$

Question c) Using the property of the Gaussian vector y, we have

$$\ln\left(\mathbb{P}\left(oldsymbol{y}\in\mathcal{C}
ight)
ight)\leq\min_{oldsymbol{\lambda}}\left\{S_{\mathcal{C}}(-oldsymbol{\lambda})+rac{oldsymbol{\lambda}^{\mathrm{T}}oldsymbol{\lambda}}{2}
ight\}.$$

Combining this with the result of question b) gives

$$egin{aligned} &\min_{oldsymbol{\lambda}\in\mathbb{R}^m}\left\{S_\mathcal{C}(-oldsymbol{\lambda})+rac{oldsymbol{\lambda}^{\mathrm{T}}oldsymbol{\lambda}}{2}
ight\} &=&\min_{oldsymbol{\lambda}\in\mathbb{R}^m}\left\{egin{aligned} η^{\mathrm{T}}oldsymbol{u} \midoldsymbol{A}^{\mathrm{T}}oldsymbol{u}+oldsymbol{\lambda}=0,oldsymbol{u}\geqoldsymbol{0}
ight\},\ &=&\min_{oldsymbol{\lambda}\in\mathbb{R}^m\atop \mu\in\mathbb{R}^\ell}\left\{eldsymbol{b}^{\mathrm{T}}oldsymbol{u}+rac{oldsymbol{\lambda}^{\mathrm{T}}oldsymbol{\lambda}}{2}\midoldsymbol{A}^{\mathrm{T}}oldsymbol{u}+oldsymbol{\lambda}=0,oldsymbol{u}\geqoldsymbol{0}
ight\}, \end{aligned}$$

which is the desired result.

d)i) The problem (6) is convex and its feasible set has a nonempty interior. Slater's condition holds, implying that strong duality holds between problem (6) and its dual. To obtain the latter, we write the dual function

$$d(\boldsymbol{\nu}) = \begin{cases} \min_{\boldsymbol{u} \in \mathbb{R}^{\ell}} \boldsymbol{b}^{\mathrm{T}} \boldsymbol{u} + \frac{\|\boldsymbol{A}^{\mathrm{T}} \boldsymbol{u}\|^{2}}{2} - \boldsymbol{\nu}^{\mathrm{T}} \boldsymbol{u} & \text{if } \boldsymbol{\nu} \geq \boldsymbol{0} \\ -\infty & \text{otherwise.} \end{cases}$$

For $\nu \geq 0$, the optimal solution of the minimization in u satisfies

$$\boldsymbol{b} - \boldsymbol{\nu} + \boldsymbol{A} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{u}^{*} = \boldsymbol{0} \quad \Rightarrow \quad d(\boldsymbol{\nu}) = -\frac{1}{2} \|\boldsymbol{A}^{\mathrm{T}} \boldsymbol{u}^{*}\|^{2}.$$

Letting $oldsymbol{x} = -oldsymbol{A}^{\mathrm{T}}oldsymbol{u}^{*} \in \mathbb{R}^{m}$, we get 4

$$d(\mathbf{\nu}) = -\frac{\|\mathbf{x}\|^2}{2}, \quad -\mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{\nu}$$

The dual problem $\operatorname{maximize}_{\boldsymbol{\nu}\geq \boldsymbol{0}} d(\boldsymbol{\nu})$ can be written as

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{\nu} \geq \mathbf{0}} & -\frac{\|\boldsymbol{x}\|^2}{2} \\ \text{s.t.} & -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{\nu}, \end{array}$$

that can be rewritten as a problem over x and u

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^\ell} & -\frac{\|\boldsymbol{x}\|^2}{2} \\ \text{s.t.} & -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{\nu} \\ \boldsymbol{\nu} \geq \boldsymbol{0}. \end{array}$$

Eliminating further the variable u, we finally arrive at

d)ii) The dual problem (7) can be reformulated as

Consider the equivalent reformulation

$$\min_{oldsymbol{x}\in\mathbb{R}^m} \|oldsymbol{x}\| \quad ext{s.t.} \quad oldsymbol{x}\in\mathcal{C}.$$

The optimal value of this problem corresponds to the definition of the distance between the zero vector and the set C, i.e.

$$\min_{oldsymbol{x}\in\mathbb{R}^m}\left\{ \|oldsymbol{x}\|\midoldsymbol{x}\in\mathcal{C}
ight\} = \mathrm{dist}(oldsymbol{0},\mathcal{C}).$$

As a result,

$$\min_{\boldsymbol{x}\in\mathbb{R}^m}\frac{1}{2}\|\boldsymbol{x}\|^2 \quad \text{s.t.} \quad \boldsymbol{x}\in\mathcal{C}=\frac{1}{2}\text{dist}(\boldsymbol{0},\mathcal{C})^2.$$

and the optimal value of problem (7) is given by $-\frac{1}{2}{\rm dist}(0,\mathcal{C})^2.$

⁴Recall that $\|\boldsymbol{x}\| = \| - \boldsymbol{x}\|$ for any vector \boldsymbol{x} .

d)iii) Since strong duality holds, the optimal value of problem (6) is $-\frac{1}{2}dist(0,C)^2$, as is that of problem (4). Plugging this result into inequalities (4) and (3), we obtain

$$\begin{split} &\ln\left(\mathbb{P}\left(\boldsymbol{y}\in\mathcal{C}\right)\right)\leq-\frac{1}{2}\mathrm{dist}(\boldsymbol{0},\mathcal{C})^2\\ &\mathbb{P}\left(\boldsymbol{y}\in\mathcal{C}\right)\leq\exp\left[-\frac{1}{2}\mathrm{dist}(\boldsymbol{0},\mathcal{C})^2\right] \end{split}$$

and

respectively.

NB: This is yet another concentration inequality.

Solution for Exercise 3.5: Random projections

a) For any column p_j of P, the distribution of p_j is given by the joint distribution of its elements. Since those are i.i.d., the joint distribution is given by the product of the marginal distributions.

$$\forall \boldsymbol{p} \in \mathbb{R}^n, p(\boldsymbol{p}) = \left(\frac{r}{2\pi}\right)^{n/2} \exp\left(-\frac{r}{2}\sum_{i=1}^n [\boldsymbol{p}]_i^2\right)$$

b) It suffices to note that for all $oldsymbol{p} \in \mathbb{R}^n$, we have

$$-\ln(p(\boldsymbol{p})) = -\ln\left(\left(\frac{r}{2\pi}\right)^{n/2} \exp\left(-\frac{r}{2}\sum_{i=1}^{n}[\boldsymbol{p}]_{i}^{2}\right)\right)$$
$$= \frac{n}{2}\ln\left(\frac{2\pi}{r}\right) + \frac{r}{2}\sum_{i=1}^{n}[\boldsymbol{p}]_{i}^{2}$$
$$= \frac{n}{2}\ln\left(\frac{2\pi}{r}\right) + \frac{r}{2}\|\boldsymbol{p}\|^{2}.$$

Since the latter function is a convex (quadratic) function of p, it follows that p is a log-concave function.

c) i) Using

$$\mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{a}\| > (1-\epsilon)\|\boldsymbol{a}\|\right) = 1 - \mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{a}\| \le (1-\epsilon)\|\boldsymbol{a}\|\right)$$

(9) can be equivalently formulated as

$$1 - \mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{a}\| \le (1-\epsilon)\|\boldsymbol{a}\|\right) \ge 0.99 \quad \Leftrightarrow \quad \mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{a}\| \le (1-\epsilon)\|\boldsymbol{a}\|\right) \le 0.01.$$

Per (8), it suffices to guarantee that $\exp(-c r\epsilon^2) \le 0.01$ in order to establish (9). ii) We have

$$\exp\left(-c\,r\epsilon^2\right) \leq 0.01$$

$$-c\,r\,\epsilon^2 \leq \ln(0.01)$$

$$c\,r\,\epsilon^2 \geq -\ln(0.01)$$

$$r \geq \frac{-\ln(0.01)}{c\epsilon^2} = \frac{\ln(100)}{c\epsilon^2}.$$

Letting $\bar{r} = \frac{\ln(100)}{c\epsilon^2}$, it follows that (9) holds for $r \geq \bar{r}$.

iii) Suppose that we generate random matrices P_1, \ldots, P_m of size $\underline{r} \times n$ independently. By assumption, it follows that

$$\mathbb{P}\left(\|\boldsymbol{P}_{i}\boldsymbol{a}\| \geq (1-\epsilon)\|\boldsymbol{a}\|\right) \geq 0.6 \qquad \forall i = 1, \dots, m$$

Then,

$$\begin{split} \mathbb{P}\left(\exists i = 1, \dots, m, \|\boldsymbol{P}_{i}\boldsymbol{a}\| \geq (1-\epsilon)\|\boldsymbol{a}\|\right) &= 1 - \mathbb{P}\left(\forall i = 1, \dots, m, \|\boldsymbol{P}_{i}\boldsymbol{a}\| < (1-\epsilon)\|\boldsymbol{a}\|\right) \\ &= 1 - \prod_{i=1}^{m} \mathbb{P}\left(\|\boldsymbol{P}_{i}\boldsymbol{a}\| < (1-\epsilon)\|\boldsymbol{a}\|\right) \\ &= 1 - \prod_{i=1}^{m} \left(1 - \mathbb{P}\left(\|\boldsymbol{P}_{i}\boldsymbol{a}\| \geq (1-\epsilon)\|\boldsymbol{a}\|\right)\right) \\ &\geq 1 - \prod_{i=1}^{m} (1-0.6) = 1 - 0.4^{m}. \end{split}$$

In order to guarantee that the probability is greater than 0.99, it suffices to have

$$1 - 0.4^m \ge 0.99 \quad \Leftrightarrow \quad m \ge \frac{\ln(0.01)}{\ln(0.4)} \approx 5.02$$

By drawing 6 independent matrices, one is thus guaranteed that one of them satisfies the desired accuracy.

Solution for Exercise 3.6: Erdös-Rényi graphs

- a) For any j ∈ {1,...,n}, j ≠ i, let e_{ij} be the random variable representing whether the edge (i, j) is included in the graph. Then e_{ij} is a Bernoulli variable of parameter p, and d_i = ∑1≤j≤n e_{ij}, with all e_{ij}s being independent by assumption.
- b) A straightforward calculation gives

$$\mathbb{E}\left[d_i\right] = \mathbb{E}\left[\sum_{j \neq i} e_{ij}\right] = \sum_{j \neq i} \mathbb{E}\left[e_{ij}\right] = \sum_{j \neq i} p = (n-1)p = \bar{d}.$$

c) Applying (10) with N = n - 1, $\{x_j\} = \{e_{ij}\}, p_j = p, \mu = \bar{d} \text{ and } \delta = 0.1 \text{ gives}$ $\mathbb{P}\left(|d_i - \bar{d}| \ge 0.1d\right) \le 2\exp(-c\bar{d}0.1^2),$

hence (11) holds with C = 0.01c.

d) By a union bound argument, Use this inequality to find a bound on

$$\mathbb{P}\left(\exists i \in \{1, \dots, n\}, \quad |d_i - \bar{d}| \ge 0.1\bar{d}\right) \le \sum_{i=1}^n \mathbb{P}\left(|d_i - \bar{d}| \ge 0.1\bar{d}\right)$$
$$\le \sum_{i=1}^n 2\exp(-C\bar{d}) = 2n\exp(-C\bar{d})$$

where the last inequality follows from (11).

e) It suffices to note that

$$\mathbb{P}\left(\max_{1\leq i\leq n} |d_i - \bar{d}| < 0.1d\right) \ge 0.9 \quad \Leftrightarrow \quad 1 - \mathbb{P}\left(\exists i \in \{1, \dots, n\}, \quad |d_i - \bar{d}| \ge 0.1\bar{d}\right) \ge 0.9$$
$$\Leftrightarrow \quad \mathbb{P}\left(\exists i \in \{1, \dots, n\}, \quad |d_i - \bar{d}| \ge 0.1\bar{d}\right) \le 0.1.$$

According to the previous question, the latter result holds provided

$$2n \exp(-C\overline{d}) \le 0.1 \quad \Leftrightarrow \quad \overline{d} \ge \frac{\ln(20n)}{C} \Leftrightarrow p \ge \frac{\ln(20) + \ln(n)}{C(n-1)}.$$

Thus, assuming (for instance) $p \ge \frac{1+\ln(20)}{C} \frac{\ln(n)}{n-1}$ gives the desired conclusion with $\hat{c} = \frac{1+\ln(20)}{C}$.