Exercises on Chapter 3: High Probability

Mathematics of Data Science, M1 IDD

November-December 2025

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 $\mathrm{Var}[h_N] = \frac{N}{4}$.
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 $\mathrm{Var}[x+y] = \mathrm{Var}[x] + \mathrm{Var}[y]$.

b) Chebyshev's inequality states that

$$\mathbb{P}(|y - \mathbb{E}[y]| \geq t) \leq \frac{\mathrm{Var}[y]}{t^2} \qquad \forall t > 0$$

for any random variable y and any $t > 0$. 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 inequality:

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

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).

Exercise 3.2: Boosting

Suppose that we perform $2\,m$ 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 $\delta \in (0, \frac{1}{2})$. 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 ith 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) Hoeffding's inequality for bounded random variables states that for any set of variables $y_1, \ldots, y_N$ that are bounded in $[m, M]$ and any $t \geq 0$, we have

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

Using Hoeffding's inequality, show that the probability expressed in question b) is at least $1-p$ when

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

for every $p \in (0,1)$.

Exercise 3.3: Random projections (Exam 2023-2024)

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

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

a) Give a formula for the probability density function of a column of $\boldsymbol{P}$ (e.g. the first one).

b) Show that this density is log-concave.

c) Given any vector $\boldsymbol{a} \in \mathbb{R}^n$ and any tolerance $\epsilon \in (0,1]$, it can be shown (using Johnson-Lindenstrauss-type arguments) that

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

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

i) Using (1), find a sufficient condition for the bound

$$\mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{a}\| > (1-\epsilon)\|\boldsymbol{a}\|\right) \geq 0.99 \tag{2}$$

to hold.

ii) Use the condition from the previous question to determine $\bar{r}$ such that (2) holds whenever $r \geq \bar{r}$.

iii) Suppose that we cannot generate matrices using $r = \bar{r}$, but that we can generate matrices $\boldsymbol{P} \in \mathbb{R}^{\underline{r} \times n}$ with $\underline{r} < \bar{r}$ such that

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

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

Exercise 3.4: Erdös-Rényi graphs

Graphs generated using the Erdös-Rényi model $G(n, p)$ are undirected random graphs with $n \geq 2$ 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 $i \in \{1, \ldots, n\}$, 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 $\mathbb{E}[d_i] = \bar{d} := (n-1)p$.

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

$$\forall \delta \in (0,1], \qquad \mathbb{P}\left(\left|\sum_{j=1}^{N} x_j - \mu\right| \geq \delta\mu\right) \leq 2\exp(-c\mu\delta^2), \tag{3}$$

where $\mu = \mathbb{E}\left[\sum_{j=1}^{N} x_j\right]$ and $c > 0$. Use (3) to show that for any $i = 1, \ldots, n$, we have

$$\mathbb{P}\left(|d_i - \bar{d}| \geq 0.1\bar{d}\right) \leq 2\exp(-C\bar{d}) \tag{4}$$

where $C > 0$ is a universal constant.

d) Use the inequality (4) to find a bound on

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

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

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

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

Exercise 3.5: 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

$$\mathbb{P}(y \geq t) \quad \leq \quad \min_{\lambda \geq 0} \mathbb{E}\left[\exp(\lambda(y-t))\right]. \tag{5}$$

a) Proving (5) amounts to proving

$$\ln\left(\mathbb{P}(y \geq t)\right) \quad \leq \quad \min_{\lambda \geq 0} \ln\left(\mathbb{E}\left[\exp(\lambda(y-t))\right]\right). \tag{6}$$

Justify that right-hand side of (6) 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*

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

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 $\ln\left(\mathbb{E}\left[\exp(\lambda y)\right]\right) = \frac{\lambda^2}{2}$. Use this property to deduce from (5) that

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

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

Exercise 3.6: Chernoff inequalities for vectors

In this exercise, we seek a Chernoff-type bound in a vector setting. More precisely, we consider a Gaussian vector $\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}_{\mathbb{R}^n}, \boldsymbol{I}_n)$ and a nonempty polyhedral set defined by $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}\}$ 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

$$\mathbb{P}\left(\boldsymbol{y} \in \mathcal{C}\right) \le \mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} + \mu\right)\right] \tag{7}$$

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

a) Using that $\mathbb{P}\left(\boldsymbol{y} \in \mathcal{C}\right) = \mathbb{E}\left[1_{\mathcal{C}}(\boldsymbol{y})\right]$, justify that any pair $(\boldsymbol{\lambda}, \mu) \in \mathbb{R}^n \times \mathbb{R}$ satisfying $\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} + \mu\right) \ge 1_{\mathcal{C}}(\boldsymbol{y})$ for every $\boldsymbol{y} \in \mathbb{R}^n$ also satisfies (7) with $-\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} \le \mu \ \forall \boldsymbol{y} \in \mathcal{C}$.

b) By considering logarithms, show that

$$\ln\left(\mathbb{P}\left(\boldsymbol{y} \in \mathcal{C}\right)\right) \le \min_{\boldsymbol{\lambda} \in \mathbb{R}^n} \left\{ S_{\mathcal{C}}(-\boldsymbol{\lambda}) + \ln \mathbb{E}\left[e^{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y}}\right] \right\}, \tag{8}$$

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

$$S_{\mathcal{C}}(\boldsymbol{y}) = \min_{\boldsymbol{u} \in \mathbb{R}^{\ell}} \left\{ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} \middle| \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{y}, \boldsymbol{u} \ge \mathbf{0} \right\}$$

for any $\boldsymbol{y} \in \mathbb{R}^n$. Show then that the right-hand side of (8) corresponds to the optimal value of the quadratic problem

$$\begin{array}{ll} \mathrm{minimize}_{\boldsymbol{\lambda} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^{\ell}} & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{v} + \frac{\|\boldsymbol{\lambda}\|^2}{2} \\ \text{s.t.} & \boldsymbol{v} \ge \mathbf{0}, \\ & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} + \boldsymbol{\lambda} = \mathbf{0}. \end{array} \tag{9}$$

d) The problem (9) is equivalent to

$$\underset{\boldsymbol{v}\in\mathbb{R}^\ell}{\text{minimize}}\, \boldsymbol{b}^{\mathrm{T}}\boldsymbol{v} + \frac{\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{v}\|^2}{2} \quad \text{s.t.} \quad \boldsymbol{v} \geq \boldsymbol{0}, \tag{10}$$

where we reformulated the problem so as to eliminate the $\boldsymbol{\lambda}$ variables while preserving the same optimal value.

i) Using that same reformulation technique, show that the dual of problem (10) is equivalent to

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{x}\in\mathbb{R}^n} & -\frac{\|\boldsymbol{x}\|^2}{2} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}. \end{array} \tag{11}$$

ii) Justify that the optimal value of problem (11) is $-\frac{1}{2}\mathrm{dist}(\boldsymbol{0},\mathcal{C})^2$, where $\mathrm{dist}(\boldsymbol{a},\mathcal{C}) = \min_{\boldsymbol{y}\in\mathcal{C}} \|\boldsymbol{y}-\boldsymbol{a}\|$.

iii) Strong duality holds for problem (10). Using this property, provide a closed-form expression for (8) and (7).