Exercises on Chapter 3: Statistics and concentration inequalities

Mathematics of Data Science, M1 IDD

November-December 2024*

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) 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[1]:

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

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).

*Version 2, last updated November 20, 2024.

[1] To 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

$$\mathbb{P}\left(y \geq t\right) \quad \leq \quad \min_{\lambda \geq 0} \mathbb{E}\left[\exp(\lambda(y-t))\right]. \tag{1}$$

a) Proving (1) amounts to proving

$$\ln\left(\mathbb{P}\left(y \geq t\right)\right) \quad \leq \quad \min_{\lambda \geq 0} \ln\left(\mathbb{E}\left[\exp(\lambda(y-t))\right]\right). \tag{2}$$

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*

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

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 (1) that

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

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

Exercise 3.3: 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 $\delta \in (0,1)$. 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 ith run returns the wrong output, and $y_i = 0$ otherwise. Compute $\mathbb{E}\left[y_i\right]$.

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

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

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 $\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}_{\mathbb{R}^n}, \boldsymbol{I}_n)$ and a nonempty polyhedral set defined by $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ 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

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

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 $\exp\left(\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} + \mu\right) \geq 1_{\mathcal{C}}(\boldsymbol{y})$ for every $\boldsymbol{y} \in \mathbb{R}^n$ also satisfies (3) with $-\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} \leq \mu \,\forall \boldsymbol{y} \in \mathcal{C}$.

b) By considering logarithms, show that

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

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

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

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

$$\begin{array}{ll} \operatorname{minimize}_{\boldsymbol{\lambda} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^{\ell}} & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{v} + \frac{\|\boldsymbol{\lambda}\|^2}{2} \\ \text{s.t.} & \boldsymbol{v} \geq \mathbf{0}, \\ & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} + \boldsymbol{\lambda} = \mathbf{0}. \end{array} \tag{5}$$

d) The problem (5) is equivalent to

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

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 (6) is equivalent to

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

   ii) Justify that the optimal value of problem (7) is $-\frac{1}{2}\mathrm{dist}(\mathbf{0}, \mathcal{C})^2$, where $\mathrm{dist}(\boldsymbol{a}, \mathcal{C}) = \min_{\boldsymbol{y} \in \mathcal{C}} \|\boldsymbol{y} - \boldsymbol{a}\|$.

   iii) Strong duality holds for problem (6). Using this property, provide a closed-form expression for (4) and (3).