Appendix

In this appendix we collect some of the technical tools used in the book and not proved in the main text. Most of the results reproduced here are quite standard; they are here to make the book as self-contained as possible. Here we take a minimalist approach and stick to the simplest possible versions that are necessary to follow the material in the main text. This appendix should not be taken as an attempt to an exhaustive survey. The cited references merely intend to point to the original source of the results.

A.1 Inequalities from Probability Theory

A.1.1 Hoeffding's Inequality

First we offer a proof of Lemma 2.2, which states the following:

***Lemma* A.1.** *Let X be a random variable with $a \le X \le b$. Then for any $s \in \mathbb{R}$,*

$$\ln \mathbb{E}\left[e^{sX}\right] \le s\,\mathbb{E}\,X + \frac{s^2(b-a)^2}{8}.$$

Proof. Since $\ln \mathbb{E}\left[e^{sX}\right] \le s\,\mathbb{E}\,X + \ln \mathbb{E}\left[e^{s(X-\mathbb{E}X)}\right]$, it suffices to show that for any random variable X with $\mathbb{E}\,X = 0$, $a \le X \le b$,

$$\mathbb{E}\left[e^{sX}\right] \le e^{s^2(b-a)^2/8}.$$

Note that by convexity of the exponential function,

$$e^{sx} \le \frac{x-a}{b-a}e^{sb} + \frac{b-x}{b-a}e^{sa} \qquad \text{for } a \le x \le b.$$

Exploiting $\mathbb{E}\,X = 0$, and introducing the notation $p = -a/(b-a)$, we get

$$\begin{aligned} \mathbb{E}e^{sX} &\le \frac{b}{b-a}e^{sa} - \frac{a}{b-a}e^{sb} \\ &= \left(1 - p + pe^{s(b-a)}\right)e^{-ps(b-a)} \\ &\stackrel{\text{def}}{=} e^{\phi(u)}, \end{aligned}$$

where $u = s(b-a)$, and $\phi(u) = -pu + \log(1-p+pe^u)$. But by straightforward calculation it is easy to see that the derivative of ϕ is

$$\phi'(u) = -p + \frac{p}{p + (1-p)e^{-u}}$$

and therefore $\phi(0) = \phi'(0) = 0$. Moreover,

$$\phi''(u) = \frac{p(1-p)e^{-u}}{(p + (1-p)e^{-u})^2} \le \frac{1}{4}.$$

Thus, by Taylor's theorem,

$$\phi(u) = \phi(0) + u\phi'(0) + \frac{u^2}{2}\phi''(\theta) \le \frac{u^2}{8} = \frac{s^2(b-a)^2}{8}$$

for some $\theta \in [0, u]$. ■

Lemma A.1 was originally proven to derive the following result, also known as *Hoeffding's inequality*.

***Corollary* A.1.** *Let $X_1, \ldots, X_n$ be independent real-valued random variables such that for each $i = 1, \ldots, n$ there exist some $a_i \le b_i$ such that $\mathbb{P}[a_i \le X_i \le b_i] = 1$. Then for every $\varepsilon > 0$,*

$$\mathbb{P}\left[\sum_{i=1}^n X_i - \mathbb{E}\sum_{i=1}^n X_i > \varepsilon\right] \le \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

and

$$\mathbb{P}\left[\sum_{i=1}^n X_i - \mathbb{E}\sum_{i=1}^n X_i < -\varepsilon\right] \le \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Proof. The proof is based on a clever application of Markov's inequality, often referred to as *Chernoff's technique*: for any $s > 0$,

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^n (X_i - \mathbb{E}X_i) > t\right] &\le \frac{\mathbb{E}\left[\exp\left(s\sum_{i=1}^n (X_i - \mathbb{E}X_i)\right)\right]}{\exp(st)} \\ &= \frac{\prod_{i=1}^n \mathbb{E}\left[\exp\left(s(X_i - \mathbb{E}X_i)\right)\right]}{\exp(st)},\end{aligned}$$

where we used independence of the variables X_i. Bound the numerator using Lemma A.1 and minimize the obtained bound in s to get the first inequality. The second is obtained by symmetry. ■

We close this section by a version of Corollary A.1, also due to Hoeffding [161], for the case when sampling is done without replacement.