# A

# Appendix



## A.1 Mathematical Symbols

$\{e_0, \ldots, e_{n-1}\}$: set containing elements $e_0$, …, $e_{n-1}$.

$\{e : P(e)\}$: set of all elements that fulfill the predicate $P$.

$\langle e_0, \ldots, e_{n-1} \rangle$: sequence consisting of elements $e_0$, …, $e_{n-1}$.

$\langle e \in S : P(e) \rangle$: subsequence of all elements of sequence $S$ that fulfill the predicate $P$.

$|x|$: the absolute value of $x$.

$\lfloor x \rfloor$: the largest integer $\leq x$.

$\lceil x \rceil$: the smallest integer $\geq x$.

$[a,b] := \{x \in \mathbb{R} : a \leq x \leq b\}$.

$i..j$: abbreviation for $\{i, i+1, \ldots, j\}$.

$A^B$: when $A$ and $B$ are sets, this is the set of all functions that map $B$ to $A$.

$A \times B$: the set of pairs $(a,b)$ with $a \in A$ and $b \in B$.

$\perp$: an undefined value.

$(-)\infty$: (minus) infinity.

$\forall x : P(x)$: for *all* values of $x$, the proposition $P(x)$ is true.

$\exists x : P(x)$: there *exists* a value of $x$ such that the proposition $P(x)$ is true.

$\mathbb{N}$: nonnegative integers; $\mathbb{N} = \{0, 1, 2, \ldots\}$.

$\mathbb{N}_+$: positive integers; $\mathbb{N}_+ = \{1, 2, \ldots\}$.

$\mathbb{Z}$: integers.

$\mathbb{R}$: real numbers.

$\mathbb{Q}$: rational numbers.

$|$, &, «, », $\oplus$: bitwise OR, bitwise AND, leftshift, rightshift, and exclusive OR respectively.

$\sum_{i=1}^{n} a_i = \sum_{1 \le i \le n} a_i = \sum_{i \in \{1,\ldots,n\}} a_i := a_1 + a_2 + \cdots + a_n$.

$\prod_{i=1}^{n} a_i = \prod_{1 \le i \le n} a_i = \prod_{i \in \{1,\ldots,n\}} a_i := a_1 \cdot a_2 \cdots a_n$.

$n! := \prod_{i=1}^{n} i$, the *factorial* of $n$.

$H_n := \sum_{i=1}^{n} 1/i$, the $n$-th *harmonic number* (Equation (A.12)).

$\log x$: The logarithm to base two of $x$, $\log_2 x$.

$\mu(s,t)$: the shortest-path distance from $s$ to $t$; $\mu(t) := \mu(s,t)$.

div: integer division; $m \operatorname{div} n := \lfloor m/n \rfloor$.

mod: modular arithmetic; $m \bmod n = m - n(m \operatorname{div} n)$.

$a \equiv b \pmod m$: $a$ and $b$ are congruent modulo $m$, i.e., $a + im = b$ for some integer $i$.

$\prec$: some ordering relation. In Sect. 9.2, it denotes the order in which nodes are marked during depth-first search.

1, 0: the boolean values "true" and "false".

## A.2 Mathematical Concepts

**antisymmetric**: a relation $\sim$ is *antisymmetric* if for all $a$ and $b$, $a \sim b$ and $b \sim a$ implies $a = b$.

**asymptotic notation**:

$$\begin{aligned}
O(f(n)) &:= \{g(n) : \exists c > 0 : \exists n_0 \in \mathbb{N}_+ : \forall n \ge n_0 : g(n) \le c \cdot f(n)\}. \\
\Omega(f(n)) &:= \{g(n) : \exists c > 0 : \exists n_0 \in \mathbb{N}_+ : \forall n \ge n_0 : g(n) \ge c \cdot f(n)\}. \\
\Theta(f(n)) &:= O(f(n)) \cap \Omega(f(n)). \\
o(f(n)) &:= \{g(n) : \forall c > 0 : \exists n_0 \in \mathbb{N}_+ : \forall n \ge n_0 : g(n) \le c \cdot f(n)\}. \\
\omega(f(n)) &:= \{g(n) : \forall c > 0 : \exists n_0 \in \mathbb{N}_+ : \forall n \ge n_0 : g(n) \ge c \cdot f(n)\}.
\end{aligned}$$

See also Sect. 2.1.

**concave**: a function $f$ is concave on an interval $[a,b]$ if

$$\forall x,y \in [a,b] : \forall t \in [0,1] : f(tx+(1-t)y) \geq t\,f(x)+(1-t)\,f(y).$$

**convex**: a function $f$ is convex on an interval $[a,b]$ if

$$\forall x,y \in [a,b] : \forall t \in [0,1] : f(tx+(1-t)y) \leq t\,f(x)+(1-t)\,f(y).$$

**equivalence relation**: a transitive, reflexive, symmetric relation.

**field**: a set of elements that support addition, subtraction, multiplication, and division by nonzero elements. Addition and multiplication are associative and commutative, and have neutral elements analogous to zero and one for the real numbers. The prime examples are $\mathbb{R}$, the real numbers; $\mathbb{Q}$, the rational numbers; and $\mathbb{Z}_p$, the integers modulo a prime $p$.

**iff**: abbreviation for "if and only if".

**lexicographic order**: the canonical way of extending a total order on a set of elements to tuples, strings, or sequences over that set. We have $\langle a_1,a_2,\ldots,a_k\rangle < \langle b_1,b_2,\ldots,b_k\rangle$ if and only if $a_1 < b_1$ or $a_1 = b_1$ and $\langle a_2,\ldots,a_k\rangle < \langle b_2,\ldots,b_k\rangle$.

**linear order**: a reflexive, transitive, weakly antisymmetric relation.

**median**: an element with rank $\lceil n/2\rceil$ among $n$ elements.

**multiplicative inverse**: if an object $x$ is multiplied by a *multiplicative inverse* $x^{-1}$ of $x$, we obtain $x \cdot x^{-1} = 1$ – the neutral element of multiplication. In particular, in a *field*, every element except zero (the neutral element of addition) has a unique multiplicative inverse.

**prime number**: an integer $n$, $n \geq 2$, is a prime iff there are no integers $a,b > 1$ such that $n = a \cdot b$.

**rank**: a one-to-one mapping $r : S \to 1..n$ is a ranking function for the elements of a set $S = \{e_1,\ldots,e_n\}$ if $r(x) < r(y)$ whenever $x < y$.

**reflexive**: a relation $\sim \subseteq A \times A$ is reflexive if $\forall a \in A : (a,a) \in R$.

**relation**: a set of pairs $R$. Often, we write relations as operators; for example, if $\sim$ is a relation, $a \sim b$ means $(a,b) \in \sim$.

**symmetric relation**: a relation $\sim$ is *symmetric* if for all $a$ and $b$, $a \sim b$ implies $b \sim a$.

**total order**: a reflexive, transitive, antisymmetric relation.

**transitive**: a relation $\sim$ is *transitive* if for all $a$, $b$, $c$, $a \sim b$ and $b \sim c$ imply $a \sim c$.

**weakly antisymmetric**: a relation $\leq$ is *weakly antisymmetric* if for all $a$, $b$, $a \leq b$ or $b \leq a$. If $a \leq b$ and $b \leq a$, we write $a \equiv b$. Otherwise, we write $a < b$ or $b < a$.

## A.3 Basic Probability Theory

Probability theory rests on the concept of a *sample space* $\mathcal{S}$. For example, to describe the rolls of two dice, we would use the 36-element sample space $\{1,\ldots,6\} \times \{1,\ldots,6\}$, i.e., the elements of the sample space are the pairs $(x,y)$ with $1 \le x,y \le 6$ and $x,y \in \mathbb{N}$. Generally, a sample space is any set. In this book, all sample spaces are finite. In a *random experiment*, any element of $s \in \mathcal{S}$ is chosen with some elementary *probability* $p_s$, where $\sum_{s \in \mathcal{S}} p_s = 1$. A sample space together with a probability distribution is called a *probability space*. In this book, we use *uniform probabilities* almost exclusively; in this case $p_s = p = 1/|\mathcal{S}|$. Subsets $\mathcal{E}$ of the sample space are called *events*. The probability of an *event* $\mathcal{E} \subseteq \mathcal{S}$ is the sum of the probabilities of its elements, i.e., $\text{prob}(\mathcal{E}) = |\mathcal{E}|/|\mathcal{S}|$ in the uniform case. So the probability of the event $\{(x,y) : x+y = 7\} = \{(1,6),(2,5),\ldots,(6,1)\}$ is equal to $6/36 = 1/6$, and the probability of the event $\{(x,y) : x+y \ge 8\}$ is equal to $15/36 = 5/12$.

A *random variable* is a mapping from the sample space to the real numbers. Random variables are usually denoted by capital letters to distinguish them from plain values. For example, the random variable $X$ could give the number shown by the first die, the random variable $Y$ could give the number shown by the second die, and the random variable $S$ could give the sum of the two numbers. Formally, if $(x,y) \in \mathcal{S}$, then $X((x,y)) = x$, $Y((x,y)) = y$, and $S((x,y)) = x+y = X((x,y)) + Y((x,y))$.

We can define new random variables as expressions involving other random variables and ordinary values. For example, if $V$ and $W$ are random variables, then $(V+W)(s) = V(s)+W(s)$, $(V \cdot W)(s) = V(s) \cdot W(s)$, and $(V+3)(s) = V(s)+3$.

Events are often specified by predicates involving random variables. For example, $X \le 2$ denotes the event $\{(1,y),(2,y) : 1 \le y \le 6\}$, and hence $\text{prob}(X \le 2) = 1/3$. Similarly, $\text{prob}(X+Y = 11) = \text{prob}(\{(5,6),(6,5)\}) = 1/18$.

*Indicator random variables* are random variables that take only the values zero and one. Indicator variables are an extremely useful tool for the probabilistic analysis of algorithms because they allow us to encode the behavior of complex algorithms into simple mathematical objects. We frequently use the letters $I$ and $J$ for indicator variables.

The *expected value* of a random variable $Z : \mathcal{S} \to \mathbb{R}$ is

$$\text{E}[Z] = \sum_{s \in \mathcal{S}} p_s \cdot Z(s) = \sum_{z \in \mathbb{R}} z \cdot \text{prob}(Z = z) \,, \tag{A.1}$$

i.e., every sample $s$ contributes the value of $Z$ at $s$ times its probability. Alternatively, we can group all $s$ with $Z(s) = z$ into the event $Z = z$ and then sum over the $z \in \mathbb{R}$.

In our example, $\text{E}[X] = (1+2+3+4+5+6)/6 = 21/6 = 3.5$, i.e., the expected value of the first die is 3.5. Of course, the expected value of the second die is also 3.5. For an indicator random variable $I$, we have

$$\text{E}[I] = 0 \cdot \text{prob}(I = 0) + 1 \cdot \text{prob}(I = 1) = \text{prob}(I = 1) \,.$$

Often, we are interested in the expectation of a random variable that is defined in terms of other random variables. This is easy for sums owing to the *linearity of*

*expectations* of random variables: for any two random variables $V$ and $W$,

$$\mathrm{E}[V+W] = \mathrm{E}[V] + \mathrm{E}[W] \; . \tag{A.2}$$

This equation is easy to prove and extremely useful. Let us prove it. It amounts essentially to an application of the distributive law of arithmetic. We have

$$\begin{aligned} \mathrm{E}[V+W] &= \sum_{s\in\mathscr{S}} p_s \cdot (V(s)+W(s)) \\ &= \sum_{s\in\mathscr{S}} p_s \cdot V(s) + \sum_{s\in\mathscr{S}} p_s \cdot W(s) \\ &= \mathrm{E}[V] + \mathrm{E}[W] \; . \end{aligned}$$

As our first application, let us compute the expected sum of two dice. We have

$$\mathrm{E}[S] = \mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] = 3.5 + 3.5 = 7 \; .$$

Observe that we obtain the result with almost no computation. Without knowing about the linearity of expectations, we would have to go through a tedious calculation:

$$\begin{aligned} \mathrm{E}[S] &= 2\cdot\tfrac{1}{36} + 3\cdot\tfrac{2}{36} + 4\cdot\tfrac{3}{36} + 5\cdot\tfrac{4}{36} + 6\cdot\tfrac{5}{36} + 7\cdot\tfrac{6}{36} + 8\cdot\tfrac{5}{36} + 9\cdot\tfrac{4}{36} + \ldots + 12\cdot\tfrac{1}{36} \\ &= \frac{2\cdot 1 + 3\cdot 2 + 4\cdot 3 + 5\cdot 4 + 6\cdot 5 + 7\cdot 6 + 8\cdot 5 + \ldots + 12\cdot 1}{36} = 7 \; . \end{aligned}$$

**Exercise A.1.** What is the expected sum of three dice?

We shall now give another example with a more complex sample space. We consider the experiment of throwing $n$ balls into $m$ bins. The balls are thrown at random and distinct balls do not influence each other. Formally, our sample space is the set of all functions $f$ from $1..n$ to $1..m$. This sample space has size $m^n$, and $f(i)$, $1 \le i \le n$, indicates the bin into which the ball $i$ is thrown. All elements of the sample space are equally likely. How many balls should we expect in bin 1? We use $I$ to denote the number of balls in bin 1. To determine $\mathrm{E}[I]$, we introduce indicator variables $I_i$, $1 \le i \le n$. The variable $I_i$ is 1, if ball $i$ is thrown into bin 1, and is 0 otherwise. Formally, $I_i(f) = 0$ iff $f(i) \neq 1$. Then $I = \sum_i I_i$. We have

$$\mathrm{E}[I] = \mathrm{E}[\sum_i I_i] = \sum_i \mathrm{E}[I_i] = \sum_i \mathrm{prob}(I_i = 1) \; ,$$

where the first equality is the linearity of expectations and the second equality follows from the $I_i$'s being indicator variables. It remains to determine the probability that $I_i = 1$. Since the balls are thrown at random, ball $i$ ends up in any bin[1] with the same probability. Thus $\mathrm{prob}(I_i = 1) = 1/m$, and hence

$$\mathrm{E}[I] = \sum_i \mathrm{prob}(I_i = 1) = \sum_i \frac{1}{m} = \frac{n}{m} \; .$$

[1] Formally, there are exactly $m^{n-1}$ functions $f$ with $f(i) = 1$.

Products of random variables behave differently. In general, we have $\mathrm{E}[X \cdot Y] \neq \mathrm{E}[X] \cdot \mathrm{E}[Y]$. There is one important exception: if $X$ and $Y$ are *independent*, equality holds. Random variables $X_1, \ldots, X_k$ are independent if and only if

$$\forall x_1, \ldots, x_k : \mathrm{prob}(X_1 = x_1 \wedge \cdots \wedge X_k = x_k) = \prod_{1 \leq i \leq k} \mathrm{prob}(X_i = x_i) \; . \tag{A.3}$$

As an example, when we roll two dice, the value of the first die and the value of the second die are independent random variables. However, the value of the first die and the sum of the two dice are not independent random variables.

**Exercise A.2.** Let $I$ and $J$ be independent indicator variables and let $X = (I + J) \bmod 2$, i.e., $X$ is one iff $I$ and $J$ are different. Show that $I$ and $X$ are independent, but that $I$, $J$, and $X$ are dependent.

Assume now that $X$ and $Y$ are independent. Then

$$\begin{aligned}
\mathrm{E}[X] \cdot \mathrm{E}[Y] &= \left( \sum_x x \cdot \mathrm{prob}(X = x) \right) \cdot \left( \sum_y y \cdot \mathrm{prob}(X = y) \right) \\
&= \sum_{x,y} x \cdot y \cdot \mathrm{prob}(X = x) \cdot \mathrm{prob}(X = y) \\
&= \sum_{x,y} x \cdot y \cdot \mathrm{prob}(X = x \wedge Y = y) \\
&= \sum_z \sum_{x,y \text{ with } z = x \cdot y} z \cdot \mathrm{prob}(X = x \wedge Y = y) \\
&= \sum_z z \cdot \sum_{x,y \text{ with } z = x \cdot y} \mathrm{prob}(X = x \wedge Y = y) \\
&= \sum_z z \cdot \mathrm{prob}(X \cdot Y = z) \\
&= \mathrm{E}[X \cdot Y] \; .
\end{aligned}$$

How likely is it that a random variable will deviate substantially from its expected value? *Markov's inequality* gives a useful bound. Let $X$ be a nonnegative random variable and let $c$ be any constant. Then

$$\mathrm{prob}(X \geq c \cdot \mathrm{E}[X]) \leq \frac{1}{c} \; . \tag{A.4}$$

The proof is simple. We have

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{z \in \mathbb{R}} z \cdot \mathrm{prob}(X = z) \\
&\geq \sum_{z \geq c \cdot \mathrm{E}[X]} z \cdot \mathrm{prob}(X = z) \\
&\geq c \cdot \mathrm{E}[X] \cdot \mathrm{prob}(X \geq c \cdot \mathrm{E}[X]) \; ,
\end{aligned}$$

where the first inequality follows from the fact that we sum over a subset of the possible values and $X$ is nonnegative, and the second inequality follows from the fact that the sum in the second line ranges only over $z$ such that $z \geq c\mathrm{E}[X]$.

Much tighter bounds are possible for some special cases of random variables. The following situation arises several times, in the book. We have a sum $X = X_1 + \cdots + X_n$ of $n$ independent indicator random variables $X_1, \ldots, X_n$ and want to bound the probability that $X$ deviates substantially from its expected value. In this situation, the following variant of the *Chernoff bound* is useful. For any $\varepsilon > 0$, we have

$$\mathrm{prob}(X < (1-\varepsilon)\mathrm{E}[X]) \leq e^{-\varepsilon^2 \mathrm{E}[X]/2}\,, \tag{A.5}$$

$$\mathrm{prob}(X > (1+\varepsilon)\mathrm{E}[X]) \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\mathrm{E}[X]}\,. \tag{A.6}$$

A bound of the form above is called a *tail bound* because it estimates the "tail" of the probability distribution, i.e., the part for which $X$ deviates considerably from its expected value.

Let us see an example. If we throw $n$ coins and let $X_i$ be the indicator variable for the $i$-th coin coming up heads, $X = X_1 + \cdots + X_n$ is the total number of heads. Clearly, $\mathrm{E}[X] = n/2$. The bound above tells us that $\mathrm{prob}(X \leq (1-\varepsilon)n/2) \leq e^{-\varepsilon^2 n/4}$. In particular, for $\varepsilon = 0.1$, we have $\mathrm{prob}(X \leq 0.9 \cdot n/2) \leq e^{-0.01 \cdot n/4}$. So, for $n = 10\,000$, the expected number of heads is 5 000 and the probability that the sum is less than 4 500 is smaller than $e^{-25}$, a very small number.

**Exercise A.3.** Estimate the probability that $X$ in the above example is larger than 5 050.

If the indicator random variables are independent and identically distributed with $\mathrm{prob}(X_i = 1) = p$, $X$ is *binomially distributed*, i.e.,

$$\mathrm{prob}(X = k) = \binom{n}{k} p^k (1-p)^{(n-k)}\,. \tag{A.7}$$

**Exercise A.4 (balls and bins continued).** Let, as above, $I$ denote the number of balls in bin 1. Show

$$\mathrm{prob}(I = k) = \binom{n}{k} \left(\frac{1}{m}\right)^k \left(1 - \frac{1}{m}\right)^{(n-k)}\,,$$

and then attempt to compute $\mathrm{E}[I]$ as $\sum_k \mathrm{prob}(I = k)k$.

## A.4 Useful Formulae

We shall first list some useful formulae and then prove some of them.

- A simple approximation to the factorial:

$$\left(\frac{n}{e}\right)^n \leq n! \leq n^n. \tag{A.8}$$

- Stirling's approximation to the factorial:

$$n! = \left(1+\mathrm{O}\left(\frac{1}{n}\right)\right)\sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{A.9}$$

- An approximation to the binomial coefficients:

$$\binom{n}{k} \leq \left(\frac{n\cdot e}{k}\right)^k. \tag{A.10}$$

- The sum of the first $n$ integers:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}. \tag{A.11}$$

- The harmonic numbers:

$$\ln n \leq H_n = \sum_{i=1}^{n} \frac{1}{i} \leq \ln n + 1. \tag{A.12}$$

- The geometric series:

$$\sum_{i=0}^{n-1} q^i = \frac{1-q^n}{1-q} \quad \text{for } q \neq 1 \text{ and } \sum_{i\geq 0} q^i = \frac{1}{1-q} \quad \text{for } 0 \leq q < 1. \tag{A.13}$$

$$\sum_{i\geq 0} 2^{-i} = 2 \quad \text{and} \quad \sum_{i\geq 0} i\cdot 2^{-i} = \sum_{i\geq 1} i\cdot 2^{-i} = 2. \tag{A.14}$$

- Jensen's inequality:

$$\sum_{i=1}^{n} f(x_i) \leq n\cdot f\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \tag{A.15}$$

for any concave function $f$. Similarly, for any convex function $f$,

$$\sum_{i=1}^{n} f(x_i) \geq n\cdot f\left(\frac{\sum_{i=1}^{n} x_i}{n}\right). \tag{A.16}$$

### A.4.1 Proofs

For (A.8), we first observe that $n! = n(n-1)\cdots 1 \le n^n$. Also, for all $i \ge 2$, $\ln i \ge \int_{i-1}^{i} \ln x\,dx$, and therefore

$$\ln n! = \sum_{2\le i\le n} \ln i \ge \int_1^n \ln x\,dx = \Big[x(\ln x - 1)\Big]_{x=1}^{x=n} \ge n(\ln n - 1)\,.$$

Thus

$$n! \ge e^{n(\ln n-1)} = (e^{\ln n-1})^n = \Big(\frac{n}{e}\Big)^n\,.$$

Equation (A.10) follows almost immediately from (A.8). We have

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \le \frac{n^k}{(k/e)^k} = \Big(\frac{n\cdot e}{k}\Big)^k\,.$$

Equation (A.11) can be computed by a simple trick:

$$\begin{aligned} 1+2+\ldots+n &= \frac{1}{2}\left((1+2+\ldots+n-1+n)+(n+n-1+\ldots+2+1)\right)\\ &= \frac{1}{2}\left((n+1)+(2+n-1)+\ldots+(n-1+2)+(n+1)\right)\\ &= \frac{n(n+1)}{2}\,. \end{aligned}$$

The sums of higher powers are estimated easily; exact summation formulae are also available. For example, $\int_{i-1}^{i} x^2\,dx \le i^2 \le \int_i^{i+1} x^2\,dx$, and hence

$$\sum_{1\le i\le n} i^2 \le \int_1^{n+1} x^2\,dx = \Big[\frac{x^3}{3}\Big]_{x=1}^{x=n+1} = \frac{(n+1)^3-1}{3}$$

and

$$\sum_{1\le i\le n} i^2 \ge \int_0^n x^2\,dx = \Big[\frac{x^3}{3}\Big]_{x=0}^{x=n} = \frac{n^3}{3}\,.$$

For (A.12), we also use estimation by integral. We have $\int_i^{i+1}(1/x)\,dx \le 1/i \le \int_{i-1}^{i}(1/x)\,dx$, and hence

$$\ln n = \int_1^n \frac{1}{x}\,dx \le \sum_{1\le i\le n}\frac{1}{i} \le 1 + \int_1^n \frac{1}{x}\,dx = 1+\ln n\,.$$

Equation (A.13) follows from

$$(1-q)\cdot\sum_{0\le i\le n-1} q^i = \sum_{0\le i\le n-1} q^i - \sum_{1\le i\le n} q^i = 1-q^n\,.$$

Letting $n$ pass to infinity yields $\sum_{i\geq 0} q^i = 1/(1-q)$ for $0 \leq q < 1$. For $q = 1/2$, we obtain $\sum_{i\geq 0} 2^{-i} = 2$. Also,

$$\begin{aligned}\sum_{i\geq 1} i \cdot 2^{-i} &= \sum_{i\geq 1} 2^{-i} + \sum_{i\geq 2} 2^{-i} + \sum_{i\geq 3} 2^{-i} + \ldots \\ &= (1 + 1/2 + 1/4 + 1/8 + \ldots) \cdot \sum_{i\geq 1} 2^{-i} \\ &= 2 \cdot 1 = 2\,.\end{aligned}$$

For the first equality, observe that the term $2^{-i}$ occurs in exactly the first $i$ sums of the right-hand side.

Equation (A.15) can be shown by induction on $n$. For $n = 1$, there is nothing to show. So assume $n \geq 2$. Let $x^* = \sum_{1\leq i\leq n} x_i/n$ and $\bar{x} = \sum_{1\leq i\leq n-1} x_i/(n-1)$. Then $x^* = ((n-1)\bar{x} + x_n)/n$, and hence

$$\begin{aligned}\sum_{1\leq i\leq n} f(x_i) &= f(x_n) + \sum_{1\leq i\leq n-1} f(x_i) \\ &\leq f(x_n) + (n-1) \cdot f(\bar{x}) = n \cdot \left(\frac{1}{n} \cdot f(x_n) + \frac{n-1}{n} \cdot f(\bar{x})\right) \\ &\leq n \cdot f(x^*)\,,\end{aligned}$$

where the first inequality uses the induction hypothesis and the second inequality uses the definition of concavity with $x = x_n$, $y = \bar{x}$, and $t = 1/n$. The extension to convex functions is immediate, since convexity of $f$ implies concavity of $-f$.