Mathematical Terms and Identities

Thanks to Andy Nguyen and Julie Tibshirani for their advice on this handout.

This handout covers mathematical notation and identities that may be useful over the course of CS166. Feel free to refer to this handout for reference on a variety of topics. If you have any suggestions on how to improve this handout, please let us know!

Set Theory

The set $\mathbb{N}$ consists of all natural numbers. That is, $\mathbb{N} = \{ 0, 1, 2, 3, \dots \}$

The set $\mathbb{Z}$ consists of all integers: $\mathbb{Z} = \{ \dots, -3, -2, -1, 0, 1, 2, 3, \dots \}$

The set $\mathbb{R}$ consists of all real numbers.

The set $\emptyset$ is the empty set consisting of no elements.

If x belongs to set S, we write $x \in S$. If x does not belong to S, we write $x \notin S$.

The union of two sets S_1 and S_2 is denoted $S_1 \cup S_2$. Their intersection is denoted $S_1 \cap S_2$, difference is denoted $S_1 - S_2$ or $S_1 \setminus S_2$, and symmetric difference is denoted $S_1 \Delta S_2$.

If S_1 is a subset of S_2, we write $S_1 \subseteq S_2$. If S_1 is a strict subset of S_2, we denote this by $S_1 \subsetneq S_2$.

The power set of a set S (denoted $\wp(S)$) is the set of all subsets of S.

The Cartesian product of two sets S_1 and S_2 is the set $S_1 \times S_2 = \{ (a, b) \mid a \in S_1 \text{ and } b \in S_2 \}$

First-Order Logic

The negations of the basic propositional connectives are as follows:

$$\neg(\neg p) \equiv p$$

$$\neg(p \land q) \equiv \neg p \lor \neg q$$

$$\neg(p \lor q) \equiv \neg p \land \neg q$$

$$\neg(p \to q) \equiv p \land \neg q$$

$$\neg(p \leftrightarrow q) \equiv p \leftrightarrow \neg q$$

The negations of the $\exists$ and $\forall$ quantifiers are as follows:

$$\neg\forall x.\ \varphi \equiv \exists x.\ \neg\varphi$$

$$\neg\exists x.\ \varphi \equiv \forall x.\ \neg\varphi$$

The statement "iff" abbreviates "if and only if."

Summations

The sum of the first n natural numbers ($0 + 1 + 2 + \ldots + n - 1$) is given by

$$\sum_{i=0}^{n-1} i = \frac{n(n-1)}{2}$$

The sum of the first n terms of the arithmetic series a, $a + b$, $a + 2b$, …, $a + (n - 1)b$ is

$$\sum_{i=0}^{n-1} (a+ib) = a\sum_{i=0}^{n-1} 1 + b\sum_{i=0}^{n-1} i = an + \frac{bn(n-1)}{2}$$

The sum of the first n terms of the geometric series 1, r, r^2, r^3, …, r^{n-1} is given by

$$\sum_{i=0}^{n-1} r^i = \frac{r^n - 1}{r - 1}$$

As a useful special case, when $r = 2$, we have

$$\sum_{i=0}^{n-1} 2^i = 2^n - 1$$

In the case that $|r| < 1$, the sum of all infinite terms of the geometric series is given by

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$$

The following summation often arises in the analysis of algorithms: when $|r| < 1$, we have

$$\sum_{i=0}^{\infty} i\, r^i = \frac{r}{(1-r)^2}$$

Inequalities

The following identities are useful for manipulating inequalities:

If $A \le B$ and $B \le C$, then $A \le C$

If $A \le B$ and $C \ge 0$, then $CA \le CB$

If $A \le B$ and $C \le 0$, then $CA \ge CB$

If $A \le B$ and $C \le D$, then $A + C \le B + D$

If $A, B \in \mathbb{Z}$, then $A \le B$ iff $A < B + 1$

If f is any monotonically increasing function and $A \le B$, then $f(A) \le f(B)$

If f is any monotonically decreasing function and $A \le B$, then $f(A) \ge f(B)$

The following inequalities are often useful in algorithmic analysis:

$$e^x \ge 1 + x$$

$$\sqrt[n]{x_1 x_2 \ldots x_n} \le \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Floors and Ceilings

The floor function $\lfloor x \rfloor$ returns the largest integer less than or equal to x. The ceiling function $\lceil x \rceil$ returns the smallest integer greater than or equal to x. These functions obey the rules

$$\lfloor x \rfloor \le x < \lfloor x \rfloor + 1 \quad \text{and} \quad \lfloor x \rfloor \in \mathbb{Z}$$

$$\lceil x \rceil - 1 < x \le \lceil x \rceil \quad \text{and} \quad \lceil x \rceil \in \mathbb{Z}$$

Additionally, $\lfloor x + n \rfloor = \lfloor x \rfloor + n$ and $\lceil x + n \rceil = \lceil x \rceil + n$ for any $n \in \mathbb{Z}$.

Asymptotic Notation

Let $f, g : \mathbb{N} \to \mathbb{N}$. Then

$$f(n) = \mathrm{O}(g(n)) \quad \text{if} \quad \exists n_0 \in \mathbb{N}.\ \exists c \in \mathbb{R}.\ \forall n \in \mathbb{N}.\ (n \ge n_0 \to f(n) \le cg(n))$$

$$f(n) = \Omega(g(n)) \quad \text{if} \quad \exists n_0 \in \mathbb{N}.\ \exists c > 0 \in \mathbb{R}.\ \forall n \in \mathbb{N}.\ (n \ge n_0 \to f(n) \ge cg(n))$$

$$f(n) = \Theta(g(n)) \quad \text{if} \quad f(n) = \mathrm{O}(g(n)) \land f(n) = \Omega(g(n))$$

When multiple variables are involved in an expression, big-O notation generalizes as follows: we say that $f(x_1, \ldots, x_n) = \mathrm{O}(g(x_1, \ldots, x_n))$ if there are constants N and c such that for any $x_1 \ge N$, $x_2 \ge N$, $\ldots$, $x_n \ge N$, we have $f(x_1, \ldots, x_n) \le c \cdot \mathrm{g}(x_1, \ldots, x_n)$.

The following rules apply for O notation:

$$\text{If } f(n) = \mathrm{O}(g(n)) \text{ and } g(n) = \mathrm{O}(h(n)), \text{ then } f(n) = \mathrm{O}(h(n)) \quad (\text{also } \Omega, \Theta, o, \omega)$$

$$\text{If } f_1(n) = \mathrm{O}(g(n)) \text{ and } f_2(n) = \mathrm{O}(g(n)), \text{ then } f_1(n) + f_2(n) = \mathrm{O}(g(n)) \quad (\text{also } \Omega, \Theta, o, \omega)$$

$$\text{If } f_1(n) = \mathrm{O}(g_1(n)) \text{ and } f_2(n) = \mathrm{O}(g_2(n)), \text{ then } f_1(n)f_2(n) = \mathrm{O}(g_1(n)g_2(n)) \quad (\text{also } \Omega, \Theta, o, \omega)$$

We can use o and ω notations to denote strict bounds on growth rates:

$$f(n) = o(g(n)) \quad \text{if} \quad \lim_{n \to \infty} \frac{f(n)}{g(n)} = 0 \qquad\qquad f(n) = \omega(g(n)) \quad \text{if} \quad \lim_{n \to \infty} \frac{f(n)}{g(n)} = \infty$$

Polynomials, exponents, and logarithms are related as follows:

$$\log_a n = \Theta(\log_b n) \text{ for any fixed constants } a, b > 1$$

$$\text{Any polynomial of degree } k \text{ with positive leading coefficient is } \Theta(n^k)$$

$$\log_b n = o(n^k) \text{ for any } k > 0$$

$$n^k = o(b^n) \text{ for any } b > 1$$

$$b^n = o(c^n) \text{ for any } 1 < b < c$$

In a graph, n denotes the number of nodes ($|V|$) and m denotes the number of edges ($|E|$). In any graph, $m = \mathrm{O}(n^2)$. In a dense graph, $m = \Theta(n^2)$; a sparse graph is one where $m = o(n^2)$.

The Master Theorem

If a, b, and d are constants, then the recurrence relation

$$\mathrm{T}(n) = a\mathrm{T}(n / b) + \mathrm{O}(n^d)$$

solves as follows:

$$\mathrm{T}(n)=\begin{cases}\mathrm{O}(n^d) & \text{if } \log_b a < d \\ \mathrm{O}(n^d \log n) & \text{if } \log_b a = d \\ \mathrm{O}(n^{\log_b a}) & \text{if } \log_b a > d\end{cases}$$

Logarithms and Exponents

Logarithms and exponents are inverses of one another: $b^{\log_b x}=\log_b b^x=x$

The change-of-base formula for logarithms states that

$$\log_b a=\frac{\log_c a}{\log_c b}$$

Sums and differences of logarithms translate into logarithms of products and quotients:

$$\log_b xy=\log_b x+\log_b y \qquad \log_b(x/y)=\log_b x-\log_b y$$

The power rule for logarithms states

$$\log_b x^y=y\log_b x$$

In some cases, exponents may be interchanged:

$$(a^b)^c=a^{bc}=(a^c)^b$$

We can change the base of an exponent using the fact that logarithms and exponents are inverses:

$$a^c=b^{c\log_b a}$$

Probability

If E_1 and E_2 are mutually exclusive events, then

$$P(E_1) + P(E_2) = P(E_1 \cup E_2)$$

For any events E_1, E_2, E_3, …, including overlapping events, the union bound states that

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} P(E_i)$$

The probability of E given F is denoted $P(E \mid F)$ and is given by

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

The chain rule for conditional probability is

$$P(E_n \cap E_{n-1} \cap \ldots \cap E_1) = P(E_n \mid E_{n-1} \cap \ldots \cap E_1) \cdot P(E_{n-1} \mid E_{n-2} \cap \ldots \cap E_1) \cdot \ldots \cdot P(E_1)$$

Two events E_1 and E_2 are called independent if

$$P(E_1 \cap E_2) = P(E_1)\, P(E_2)$$

For any event E, the complement of that event (denoted $\overline{E}$) represents the event that E does not occur. E and $\overline{E}$ are mutually exclusive, and

$$P(E) + P(\overline{E}) = 1$$

Expected Value

The expected value of a discrete random variable X is defined as

$$\mathrm{E}[X] = \sum_{i=0}^{\infty} (i\, P(X=i))$$

The expected value operator is linear: for any $a, b \in \mathbb{R}$ and any random variable X:

$$\mathrm{E}[aX + b] = a\mathrm{E}[X] + b$$

More generally, if X_1, X_2, X_3, … X_n are any random variables, then

$$\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i]$$

If X and Y are independent random variables, then

$$\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$$

Variance and Covariance

The variance of a random variable X is defined as

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2]$$

Equivalently:

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$

Given two random variables X and Y, the covariance of X and Y is defined as

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

Equivalently:

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

Accordingly:

$$\mathrm{Var}[X] = \mathrm{Cov}[X, X]$$

Variance is not a linear operator:

$$\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + 2ab\,\mathrm{Cov}[X, Y] + b^2\mathrm{Var}[Y]$$

The variance of a summation of random variables, including dependent variables, can be simplified using the following rule:

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j]$$

If X and Y are independent random variables, then $\mathrm{Cov}[X, Y] = 0$. However, if $\mathrm{Cov}[X, Y] = 0$, it is not necessarily the case that X and Y are independent.

Concentration Inequalities

Markov's inequality states that for any nonnegative random variable X with finite expected value μ and any $c > 0$, we have

$$P(X \geq c\mu) \leq \frac{1}{c}$$

Chebyshev's inequality states that for any random variable X with finite expected value μ and variance σ^2, that

$$P(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2}$$

The Chernoff bound says that if $X \sim \mathrm{Binom}(n, p)$ for $p < ½$, that

$$P\left(X \geq \frac{n}{2}\right) \leq e^{\frac{-n(1/2 - p)^2}{2p}}$$

In the case where p is a fixed constant, notice that the right-hand side is $e^{-O(1) \cdot n}$.

Useful Probability Equalities and Inequalities

An indicator random variable is a random variable X where

$$X = \begin{cases} 1 & \text{if event } F \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

For any indicator variable, $E[X] = P(F)$. Indicator variables are Bernoulli random variables, so if X is an indicator variable, then $\mathrm{Var}[X] = P(F)(1 - P(F))$.

If $X_1, X_2, \ldots, X_n$ are random variables, then

$$P(\max\{X_1, X_2, \ldots, X_n\} \le k) = P(X_1 \le k \cap X_2 \le k \cap \ldots X_n \le k)$$

$$P(\min\{X_1, X_2, \ldots, X_n\} \ge k) = P(X_1 \ge k \cap X_2 \ge k \cap \ldots X_n \ge k)$$

On expectation, repeatedly flipping a biased coin that comes up heads with probability p requires $1/p$ trials before the coin will come up heads.

Harmonic Numbers

The nth harmonic number, denoted H_n, is given by

$$H_n = \sum_{i=1}^{n} \frac{1}{i}$$

The harmonic numbers are close in value to $\ln n$: for any $n \ge 1$, we have

$$\ln(n+1) \le H_n \le \ln n + 1,$$

so $H_n = \Theta(\log n)$