Useful Inequalities (for the whole semester)

$$
\begin{aligned}
1 + x &\leq e^x, && \forall x \in \mathbb{R}, \\
1 + x &\to e^x, && \text{as } x \to 0, \\
(1+a)^n &\leq e^{an}, && \text{for } a > -1,\ n \geq 0, \\
\left(1 + \tfrac{a}{n}\right)^{bn} &\to e^{ab}, && \text{as } n \to \infty, \\
\left(\tfrac{n}{k}\right)^k \leq \binom{n}{k} &\leq \tfrac{n^k}{k!} \leq \left(\tfrac{en}{k}\right)^k, && \text{for } 1 \leq k \leq n, \\
n! &\geq \tfrac{n^n}{e^n}, && \text{for } n \geq 1, \\
n! &\sim \tfrac{n^n}{e^n}\sqrt{2\pi n}, && \text{as } n \to \infty.
\end{aligned}
$$

Probability primer

The following definitions are likely more formal than you will need, so don't worry if you have never seen the word "sigma-algebra". There is a list of important take-aways after the formal definitions.
A *sample space* is a set of elements that we call *outcomes*. An *event* is a subset $A \subseteq \Omega$. The *event space* $\mathcal{F}$ is a sigma-algebra: a collection of subsets $\mathcal{F} \subseteq 2^\Omega$ that includes Ω, is closed under taking complements and countable unions. For our purposes it is fine to assume $\mathcal{F} = 2^\Omega$. A *probability measure* is a function $\mathbb{P} : \mathcal{F} \to [0,1]$, such that $\mathbb{P}(\Omega) = 1$, and $\mathbb{P}$ is countable additive, meaning that $\mathbb{P}(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mathbb{P}(A_i)$, for disjoint events $A_1, A_2, \ldots$. Now, a *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$.

A *random variable* is a measurable function $X : \Omega \to S$, where S is the *state space* of X. We will always deal with $S \subseteq \mathbb{R}$, and most of the time we will deal with discrete nonnegative random variables, such that $S \subseteq \mathbb{N}_0$. We use the notation $\mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\})$. (Note that in that expression $B \subseteq \mathbb{R}$.) We let I_A be an *indicator random variable* for the event A. This means that

$$
I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{otherwise.} \end{cases}
$$

In the case of a discrete random variable, we define the *expectation* of X as

$$
\mathbb{E}(X) = \sum_{x \in S} x \mathbb{P}(X = x).
$$

We say that two events A_1 and A_2 are *independent* if

$$
\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).
$$

In order to understand this idea more intuitively, we need the following definition. We let

$$\mathbb{P}(A_1|A_2) = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_2)}.$$

The *conditional probability* $\mathbb{P}(A_1|A_2)$ indicates the probability of event A_1 under the assumption of A_2. Another way of thinking about conditional probabilities is the following: we assume that $\mathbb{P}(A_2) = 1$ and therefore $\mathbb{P}(\omega) = 0$ for any $\omega \notin A_2$. We then obtain a new probability measure by scaling $\mathbb{P}(\omega)$ by $1/\mathbb{P}(A_2)$ for each $\omega \in A_2$ so that our total probability remains 1. Now, we see that two events A_1 and A_2 are independent if and only if

$$\mathbb{P}(A_1|A_2) = \mathbb{P}(A_1).$$

In other words, the events A_1 and A_2 are independent if knowledge of one event does not yield information about the probability of the other.
Similarly, for random variables X, Y, we say that X and Y are *independent* if

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)(Y = y), \text{ for all } x, y.$$

Exercise 1. *Write a short overview for yourself of the above definitions: which objects are sets, which are functions?* 1

Exercise 2. *Two sets A_1 and A_2 may be disjoint (or not) and they may be independent (or not). Note that these are different, but often confused, concepts. How do they relate to one another?* 1

The following bound is often useful when we want to bound the probability of at least one event in a set of events that all have small probability, and where we do not know their interactions.

Lemma 1. *[Union bound]* $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Exercise 3. *Prove Lemma 1.* 1

We will use the following Lemma very often; it is one of the reasons that first-moment method proofs are usually surprisingly straight-forward.

Lemma 2. *We have* linearity of expectation*:*

$$\mathbb{E}\left(\sum_i X_i\right) = \sum_i \mathbb{E}(X_i),$$

whether the events X_i are dependent or not.

Exercise 4. *Prove Lemma 2.* 1

Before we get into more complicated first-moment lemmas, we make a simple observation:

Proposition 3. *We have*

$$\mathbb{P}(X \leq \mathbb{E}(X)) > 0,$$
$$\mathbb{P}(X \geq \mathbb{E}(X)) > 0.$$

Exercise 5. *Prove Proposition 3.* 1

This simple observation gives us a powerful method of proving the existence of structures without finding explicit constructions. If $X(G)$ is an invariant of, for example, graphs, and we sample graphs from some distribution, then we immediately know that graphs with $X \geq \mathbb{E}(X)$ as well as graphs with $X \leq \mathbb{E}(X)$ exist.

Lemma 4 (Markov's inequality)**.** *For a non-negative random variable, we have*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Proof. We have

$$X \;=\; X \cdot I_{X \geq t} + X \cdot I_{X<t} \;\geq\; X \cdot I_{X \geq t} \;\geq\; t \cdot I_{X \geq t}.$$

Since $\mathbb{E}(I_A) = \mathbb{P}(A)$, we have

$$\mathbb{E}(X) \geq \mathbb{E}(t \cdot I_{X \geq t}) = t \cdot \mathbb{E}(I_{X \geq t}) = t \cdot \mathbb{P}(X \geq t).$$

Thus,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

□

Exercise 6. *Where in the proof of Lemma 4 did we use the fact that X is nonnegative?* 1

Lemma 5 (First moment method)**.** *For a non-negative, integer-valued random variable, we have*

$$\mathbb{P}(X > 0) \leq \mathbb{E}(X).$$

Exercise 7. *Prove Lemma 5.* 1

Application: Hypergraph coloring

A hypergraph is a generalization of a graph in which edges can join more than two vertices. More concretely, a hypergraph is a graph (V, E) in which the edge set E is a collection of any non-empty subsets of V. A k-uniform hypergraph is a hypergraph in which every $e \in E$ has $|e| = k$.

A hypergraph is 2-colorable if the vertices can be colored red/blue such that no edge is monochromatic.

Lemma 6. *All k-uniform hypergraphs on $\leq 2^{k-1}$ edges are 2-colorable, for $k \geq 2$.*

Proof. Suppose we color the vertices of a k-uniform hypergraph red/blue uniformly and i.i.d. Then the probability an edge is monochrome red is 2^{-k} and so the probability the edge is monochrome is 2^{1-k} since there are two possible colors.
There are $|E|$ edges in total, so the expected number of monochromatic edges is

$$\mathbb{E}(X) = |E| \; \mathbb{P}(e \text{ is monochrome}) = \frac{|E|}{2^{k-1}}.$$

Suppose that $|E| = 2^{k-1}$ so that $\mathbb{E}(X) = 1$. We claim that $\mathbb{P}(X = 0) > 0$ (implying that there must exist a coloring where no edge is monochrome). Suppose that $\mathbb{P}(X = 0) = 0$. Because X is a non-negative integer valued random variable, this implies that X is constant and equal to 1. We know that this is not true, because there is an obvious coloring that achieves more than one monochrome edge (color all vertices blue). Therefore, $\mathbb{P}(X = 0) = 0$ is not possible. We conclude that all k-uniform hypergraphs on $\leq 2^{k-1}$ edges are 2-colorable. □

Application: Ramsey Theory

Let $G = (V, E)$ be a graph, and let $n = |V|$ be the number of vertices in G. A k-edge-coloring of a graph is a coloring that uses k colors. We wish to find the largest n s.t. there exists a 2-edge coloring such that G contains no monochromatic K_s.

Exercise 8. *Show that $R(s, s) > n$ if $\binom{n}{s}2^{1-\binom{s}{2}} \leq 1$.* 1

Exercise 9. *Use the inequalities at the start of the document to prove that $R(s, s) > \lfloor 2^{s/2} \rfloor$.* 2

Application: edge cuts

We let $G = (V, E)$ be a graph, with vertex set V and edge set $E \in \binom{V}{2}$. For a subset $S \subseteq V$, we let $[S, \overline{S}]$ denote the set of edges that have one endpoint in S and one endpoint in $\overline{S} = V \setminus S$. We use the probabilistic method to prove the following proposition.

Proposition 7. *For any graph G, there exists a subset $S \subseteq V(G)$ such that*

$$|[S, \overline{S}]| \geq \frac{|E|}{2}.$$

Exercise 10. *Prove Proposition 7.* 1

Exercise 11. *Can you improve on Proposition 7 by sampling S differently?* 2

Application: MAX-k-SAT

The MAX-k-SAT problem refers to the problem of determining the maximum number of clauses that can be satisfied in a Boolean formula in conjunctive normal form, where each clause has exactly k literals. A boolean formula is in *conjunctive normal form* (CNF) if it consists of a number of clauses joined by conjunctions (logical AND), and terms within the clauses are joined by disjunctions (logical OR). Example:

$$\Phi = (x_1 \vee x_2 \vee x_5) \wedge (x_3 \vee \overline{x_4} \vee \overline{x_2})$$

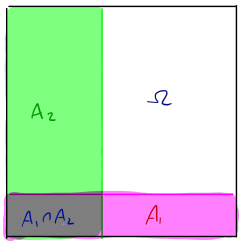
is in conjunctive normal form. Such a formula is *satisfied* by an assignment of truth values $\{0, 1\}$ to the literals makes the entire statement (Φ) true. One question we might ask is: What is the maximum number of clauses we can satisfy?

Exercise 12. *Show that in the MAX-k-SAT problem there always exists an assignment of variables that satisfies $\frac{2^k-1}{2^k}$ of clauses.* 1

Solutions to Selected Exercises

Exercise 2. *Two sets A_1 and A_2 may be disjoint (or not) and they may be independent (or not). Note that these are different, but often confused, concepts. How do they relate to one another?*

If two events A_1 and A_2 are disjoint (and assuming they both have nonzero probability), then they must be dependent, since this implies that $\mathbb{P}(A_1|A_2) = \mathbb{P}(A_2|A_1) = 0$. If two events are independent, this implies that, in a sense, A_1 is represented in A_2 as it is in Ω, and vice versa. In other words, the proportion of probability in A_2 that is also in A_1 is the same as the proportion of total probability taken up by A_1. I like to use the following sketch of independent events. (It's just a sketch and it assumes probability proportional to area.)



Exercise 3. *Prove Lemma 1.*

I may have written this on the board as a finite union instead of the infinite one in the notes, so let's prove both. The proof below for the infinite case works easily for the finite one as well, but the finite case can be proved using methods that some of you might be more familiar with.

Lemma 1 (Union bound)**.** $\mathbb{P}(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$.

Proof. First, consider $\mathbb{P}(A \cup B)$. The events, A, B may not be disjoint, but we can rewrite $A \cup B$ as a union of disjoint events as follows:

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A).$$

Then, by the properties of probability functions, we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B).$$

Furthermore, since $A = (A \setminus B) \cup (A \cap B)$ (and similarly for B), we have that

$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + 2\mathbb{P}(A \cap B).$$

Since probabilities are always nonnegative, this gives the result that

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

(Note how similar the above statement is to facts about set cardinalities that you are likely familiar with.) To generalize this to a union over n events, we can use induction. Suppose

that the result holds for up to $n-1$ events. Then, we let $B = \bigcup_{i=1}^{n-1} A_i$, to see that

$$\begin{aligned}\mathbb{P}(\bigcup_{i=1}^{n} A_i) &= \mathbb{P}(B \cup A_n) \\ &\leq \mathbb{P}(B) + \mathbb{P}(A_n) \\ &= \mathbb{P}(\bigcup_{i=1}^{n-1} A_i) + \mathbb{P}(A_n) \\ &\leq \sum_{i=1}^{n-1} \mathbb{P}(A_i) + \mathbb{P}(A_n) \\ &= \sum_{i=1}^{n} \mathbb{P}(A_i)\end{aligned}$$

Note that we use the inductive hypothesis twice here. □

Now, for the infinite case, we use a less obvious relabelling trick.

Lemma 1 (Union bound)**.** $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Proof. For each i, we let $B_i = A_i \setminus \bigcup_{j=1}^{i-1} A_j$. Note that all the events B_i are disjoint, that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$, and that $B_i \subseteq A_i$ for each i. The latter implies that $\mathbb{P}(B_i) \leq \mathbb{P}(A_i)$ for all i. Then

$$\begin{aligned}\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) &= \mathbb{P}(\bigcup_{i=1}^{\infty} B_i) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(B_i) \\ &\leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).\end{aligned}$$

□

Exercise 5. *Prove Proposition 3.*

Proposition 3. *We have*

$$\begin{aligned}\mathbb{P}(X \leq \mathbb{E}(X)) > 0, \\ \mathbb{P}(X \geq \mathbb{E}(X)) > 0.\end{aligned}$$

Proof. The two parts are very similar, so we'll only do the first one. We'll work by contradiction. Suppose (for the sake of contradiction) that $\mathbb{P}(X \leq \mathbb{E}(X)) = 0$. Let S be the set of values that X can take. Then we have $x > \mathbb{E}(X)$ for all $x \in S$. This gives

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x \in S} x\mathbb{P}(X = x) \\ &> \sum_{x \in S} \mathbb{E}(X)\mathbb{P}(X = x) \\ &= \mathbb{E}(X) \sum_{x \in S} \mathbb{P}(X = x) \\ &= \mathbb{E}(X),\end{aligned}$$

which is clearly a contradiction. □

Exercise 6. *Where in the proof of Lemma 4 did we use the fact that X is nonnegative?*

We use it in this step:

$$X \cdot I_{X \geq t} + X \cdot I_{X<t} \geq X \cdot I_{X \geq t},$$

which is only valid because

$$X \cdot I_{X<t} \geq 0.$$

Exercise 7. *Prove Lemma 5.*

Lemma 5 (First moment method)**.** *For a non-negative, integer-valued random variable, we have*

$$\mathbb{P}(X > 0) \leq \mathbb{E}(X).$$

Proof. Note that if X is non-negative and integer-valued, we have that $X > 0$ is equivalent to $X \geq 1$. Then we apply Lemma 4 with $t = 1$. □

Exercise 8. *Show that $R(s,s) > n$ if $\binom{n}{s} 2^{1-\binom{s}{2}} \leq 1$.*

Fix n and color the edges of K_n red/blue with equal probability, independently from each other. For any $R \subseteq V$ of size s let A_R to be the event that R induces a monotone K_s. The probability that R is monochrome red (or, respectively, blue) is $2^{-\binom{s}{2}}$, so the probability that R is monochrome is $2^{1-\binom{s}{2}}$. Therefore,

$$\mathbb{P}(A_R) = 2^{1-\binom{s}{2}}.$$

Let X be the number of monochrome copies of K_s in K_n. Then

$$\mathbb{E}(X) = \sum_R \mathbb{E}(I_{A_R}) = \binom{n}{s} 2^{1-\binom{s}{2}}$$

as there are $\binom{n}{s}$ possible subsets of vertices which have size s. If $\mathbb{E}(X) < 1$ then we have $\mathbb{P}(X = 0) > 0$ by the FMM, which shows that there exists a coloring with no monochromatic K_s. Therefore, $R(s,s) > n$ if

$$\binom{n}{s} 2^{1-\binom{s}{2}} \leq 1.$$

Exercise 9. *Use the inequalities at the start of the document to prove that $R(s,s) > \lfloor 2^{s/2} \rfloor$.*

Without hints, this one is very tricky! We can rewrite

$$\binom{n}{s} 2^{1-\binom{s}{2}} < \frac{n^s}{s!} \cdot \frac{2^{1+s/2}}{2^{s^2/2}} < \frac{2^{1+s/2}}{s!} \cdot \frac{n^s}{2^{s^2/2}} \leq 1.$$

Note that for $s = 2$, this question is not very interesting (clearly, $R(2,2) = 2$). When $s \geq 3$, we have that $\frac{2^{1+s/2}}{s!} < 1$. Therefore, we achieve the above inequality when $\frac{n^s}{2^{s^2/2}} \leq 1$. This works when $n = \lfloor 2^{s/2} \rfloor$.

Exercise 11. *Can you improve on Proposition 7 by sampling S differently?*

One way to do this is to note that when we distribute vertices equally over two sets, the number of pairs that span the two sets is greater than the numner of pairs within sets. Let's sample S randomly, by choosing a set S of exactly $\lfloor n/2 \rfloor$ vertices uniformly over all such sets. Let X be the size of the cut $[S, \overline{S}]$. This gives, depending on whether n is even or odd,

$$\mathbb{E}(X) = |E| \cdot \frac{\lfloor n/2 \rfloor \cdot \lceil n/2 \rceil}{\binom{n}{2}} = \begin{cases} |E| \cdot \frac{k}{2k-2}, & \text{if } n = 2k, \\ |E| \cdot \frac{k}{2k-1}, & \text{if } n = 2k-1. \end{cases}$$

In both cases this improves slightly on the bound $|E| \cdot \frac{1}{2}$.

Exercise 12. *Show that in the MAX-k-SAT problem there always exists an assignment of variables that satisfies $\frac{2^k-1}{2^k}$ of clauses.*

Consider random assignments of values to the literals such that x_i is True/False with equal probability, independently of other literals. Now consider the first clause, c_1. There are 2^k ways to assign truth values to the k literals in the clause, and there is only one of all possible assignments such that c_1 is not satisfied (when all of its literals evaluate to false). This applies the same to all clauses. Therefore, we have $\mathbb{P}(c_i) = \frac{2^k-1}{2^k}$, for all i. By linearity of expectation, this implies that the expected number of clauses that is satisfied is a proportion $\frac{2^k-1}{2^k}$ of them, and therefore there must exist some assignment that achieves at least this number.