COMP760, LECTURE 1: BASIC FUNCTIONAL ANALYSIS

HAMED HATAMI

The aim of this lecture is to introduce the necessary definitions, notations, and basic results from measure theory, and functional analysis for this course.

1. Some basic inequalities

One of the most basic inequalities in analysis concerns the arithmetic mean and the geometric mean. It is sometimes called the AM-GM inequality.

Theorem 1.1. *The geometric mean of n non-negative reals is less than or equal to their arithmetic mean: If $a_1, \ldots, a_n$ are non-negative reals, then*

$$(a_1 \ldots a_n)^{1/n} \leq \frac{a_1 + \ldots + a_n}{n}.$$

In 1906 Jensen founded the theory of convex functions. This enabled him to prove a considerable extension of the AM-GM inequality. Recall that a subset D of a real vector space is called *convex* if every convex linear combination of a pair of points of D is in D. Equivalently, if $x, y \in D$, then $tx + (1-t)y \in D$ for every $t \in [0,1]$. Given a convex set D, a function $f : D \to \mathbb{R}$ is called *convex* if for every $t \leq [0,1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

If the inequality is strict for every $t \in (0,1)$, then the function is called *strictly convex*.

Trivially f is convex if and only if $\{(x,y) \in D \times \mathbb{R} : y \geq f(x)\}$ is convex. Also note that $f : D \to \mathbb{R}$ is convex if and only if $f_{xy} : [x,y] \to \mathbb{R}$ defined as $f_{xy} : tx + (1-t)y \mapsto tf(x) + (1-t)f(y)$ is convex. By Rolle's theorem if f_{xy} is twice differentiable, then this is equivalent to $f''_{xy} \geq 0$.

A function $f : D \to \mathbb{R}$ is *concave* if $-f$ is convex. The following important inequality is often called Jensen's inequality.

Theorem 1.2. *If $f : D \to \mathbb{R}$ is a concave function, then for every $x_1, \ldots, x_n \in D$ and $t_1, \ldots, t_n \geq 0$ with $\sum_{i=1}^n t_i = 1$ we have*

$$t_1 f(x_1) + \ldots + t_n f(x_n) \leq f(t_1 x_1 + \ldots + t_n x_n).$$

Furthermore if f is strictly concave, then the equality holds if and only if all x_i are equal.

The most frequently used inequalities in functional analysis are the Cauchy-Schwarz inequality, Hölder's inequality, and Minkowski's inequality.

Theorem 1.3 (Cauchy-Schwarz)**.** *If $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are complex numbers, then*

$$\left| \sum_{i=1}^n x_i \overline{y_i} \right| \leq \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2} \left(\sum_{i=1}^n |y_i|^2 \right)^{1/2}.$$

Hölder's inequality is an important generalization of the Cauchy-Schwarz inequality.

Theorem 1.4 (Hölder's inequality)**.** *Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be complex numbers, and $p, q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\left|\sum_{i=1}^{n} x_i \overline{y_i}\right| \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}.$$

The numbers p and q appearing in Theorem 1.4 are called *conjugate exponents*. In fact 1 and ∞ are also called conjugate exponents, and Hölder's inequality in this case becomes:

$$\left|\sum_{i=1}^{n} x_i \overline{y_i}\right| \leq \left(\sum_{i=1}^{n} |x_i|\right) \left(\max_{i=1}^{n} |y_i|\right).$$

The next theorem is called Minkowski's inequality.

Theorem 1.5 (Minkowski's inequality)**.** *If $p > 1$ is a real number, and $x_1, \ldots, x_n$ are complex numbers, then*

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{1/p}.$$

The case of $p = \infty$ of Minkowski's inequality is the following:

$$\max_{i=1}^{n} |x_i + y_i| \leq \left(\max_{i=1}^{n} |x_i|\right) + \left(\max_{i=1}^{n} |y_i|\right).$$

Exercise 1.6. Let $x = \langle x_1, \ldots, x_n \rangle$ and $y = \langle y_1, \ldots, y_n \rangle$ be complex vectors. By studying the derivative of $\langle x + ty, y \rangle$ with respect to t, prove Theorem 1.3. ■

Exercise 1.7. Deduce Theorem 1.5 from Hölder's inequality. ■

2. Measure spaces

A *σ-algebra* (sometimes *sigma-algebra*) over a set Ω is a collection $\mathcal{F}$ of subsets of Ω with satisfies the following three properties:

- It includes $\emptyset$. That is, we have $\emptyset \in \mathcal{F}$.
- It is closed under complementation. That is, if $A \in \mathcal{F}$, then the complement of A also belongs to $\mathcal{F}$.
- It is closed under countable unions of its members. That is, if $A_1, A_2, \ldots$ belong to $\mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Example 2.1. Let Ω be an arbitrary set. Then the family consisting only of the empty set and the set Ω is called the *minimal* or *trivial* σ-algebra over Ω. The power set of Ω, denoted by $\mathcal{P}(\Omega)$, is the *maximal* σ-algebra over Ω. ■

There is a natural partial order between σ-algebras over Ω. For two σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ over Ω, if $\mathcal{F}_1 \subseteq \mathcal{F}_2$ then we say that $\mathcal{F}_1$ is *finer* than $\mathcal{F}_2$, or that $\mathcal{F}_2$ is *coarser* than $\mathcal{F}_1$. Note that the trivial σ-algebra is the coarsest σ-algebra over Ω, whilst the maximal σ-algebra is the finest σ-algebra over Ω.

Definition 2.2. *A* measure space *is a triple $(\Omega, \mathcal{F}, \mu)$ where $\mathcal{F}$ is a σ-algebra over Ω and the measure $\mu : \mathcal{F} \to \mathbb{R}^+ \cup \{+\infty\}$ satisfies the following axioms:*

- *Null empty set:* $\mu(\emptyset) = 0$.

- *Countable additivity: if $\{E_i\}_{i\in\mathcal{I}}$ is a countable set of pairwise disjoint sets in $\mathcal{F}$, then*
$$\mu(\cup_{i\in\mathcal{I}} E_i) = \sum_{i\in\mathcal{I}} \mu(E_i).$$

The function μ is called a measure*, and the elements of $\mathcal{F}$ are called* measurable sets*. If furthermore $\mu : \mathcal{F} \to [0,1]$ and $\mu(\Omega) = 1$, then $(\Omega, \mathcal{F}, \mu)$ is called a probability measure.*

Example 2.3. The counting measure on Ω is defined in the following way. The measure of a subset is taken to be the number of elements in the subset, if the subset is finite, and ∞ if the subset is infinite. ■

A measure space $\mathcal{M} = (\Omega, \mathcal{F}, \mu)$ is called *σ-finite*, if Ω is the countable union of measurable sets of finite measure.

$\boxed{\text{Every measure space in this this course is assumed to be } \sigma\text{-finite.}}$

For many natural measure spaces $\mathcal{M} = (\Omega, \mathcal{F}, \mu)$, it is difficult to specify the elements of the σ-algebra $\mathcal{F}$. Instead one specifies an "algebra" of elements of Ω which generates $\mathcal{F}$.

Definition 2.4. *For a set Ω, a collection $\mathcal{A}$ of subsets of Ω is called an* algebra *if*

- $\emptyset \in \mathcal{A}$.
- $A, B \in \mathcal{A}$, *then* $A \cup B \in \mathcal{A}$.
- $A, B \in \mathcal{A}$, *then* $A \setminus B \in \mathcal{A}$.

The minimal σ-algebra containing $\mathcal{A}$ is called the σ-algebra generated by $\mathcal{A}$.

Example 2.5. Let $\mathcal{A}$ be the set of all finite unions of (open, closed, or half-open) intervals in $\mathbb{R}$. Then $\mathcal{A}$ is an algebra over $\mathbb{R}$. ■

Before proceeding, let us mention that $\mu : \mathcal{A} \to \mathbb{R}^+ \cup \{+\infty\}$ is called a measure over $\mathcal{A}$ if for every finite set of $E_1, \ldots, F_m \in \mathcal{A}$, we have
$$\mu(\cup_{i=1}^m E_i) = \sum_{i=1}^m \mu(E_i).$$
The following theorem, due to Carathéodory, is one of the basic theorems in measure theory. It says that if the measure μ is defined on the algebra, then we can automatically extend it to the σ-algebra generated by $\mathcal{A}$.

Theorem 2.6 (Carathéodory's extension theorem)**.** *Let $\mathcal{A}$ be an algebra of subsets of a given set Ω. One can always extend every σ-finite measure defined on $\mathcal{A}$ to the σ-algebra generated by $\mathcal{A}$; moreover, the extension is unique.*

Example 2.7. Let $\mathcal{A}$ be the algebra on $\mathbb{R}$, defined in Example 2.5. Let μ be the measure on $\mathcal{A}$, defined by setting the measure of an interval to its length. By Carathéodory's extension theorem, μ extends uniquely to the σ-algebra generated by $\mathcal{A}$. The resulting measure is called the *Borel* measure on $\mathbb{R}$. ■

Consider two measure spaces $\mathcal{M} := (\Omega, \mathcal{F}, \mu)$ and $\mathcal{N} := (\Sigma, \mathcal{G}, \nu)$. The *product measure* $\mu \times \nu$ on $\Omega \times \Sigma$ is defined in the following way: For $F \in \mathcal{F}$ and $G \in \mathcal{G}$, define $\mu \times \nu(F \times G) = \mu(F) \times \nu(G)$. So far we defined the measure $\mu \times \nu$ on $A := \{F \times G : F \in \mathcal{F}, G \in \mathcal{G}\}$. Note that A is an algebra in that $\emptyset \in A$, and A is closed under complementation and *finite* unions of its members. However, A is not necessarily a σ-algebra, as it is possible that A is not closed under countable unions of its members. Let $\mathcal{F} \times \mathcal{G}$ be the σ-algebra generated by A, i.e. it is obtained by closing A under complementation and countable unions. It should be noted that $\mathcal{F} \times \mathcal{G}$ is *not* the cartesian product of the two sets $\mathcal{F}$ and $\mathcal{G}$, and instead it is the σ-algebra generated by the cartesian product of $\mathcal{F}$ and $\mathcal{G}$. Theorem 2.6 shows that $\mu \times \nu$ extends uniquely from A to a measure over all of $\mathcal{F} \times \mathcal{G}$. We denote the corresponding measure space by $\mathcal{M} \times \mathcal{N}$ which is called the *product measure* of $\mathcal{M}$ and $\mathcal{N}$.

Consider two measure spaces $\mathcal{M} = (\Omega, \mathcal{F}, \mu)$ and $\mathcal{N} = (\Sigma, \mathcal{G}, \nu)$. A function $f : \Omega \to \Sigma$ is called *measurable* if the preimage of every set in $\mathcal{G}$ belongs to $\mathcal{F}$.

We finish this section by stating the Borel-Cantelli theorem.

Theorem 2.8 (Borel-Cantelli)**.** *Let (E_n) be a sequence of events in some probability space. If the sum of the probabilities of the E_n is finite, then the probability that infinitely many of them occur is 0, that is,*

$$\sum_{n=1}^{\infty} \Pr[E_n] < \infty \Rightarrow \Pr[\limsup_{n\to\infty} E_n] = 0,$$

where

$$\limsup_{n\to\infty} E_n := \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{n} E_k.$$

3. Normed spaces

A *metric space* is an ordered pair (M, d) where M is a set and d is a *metric* on M, that is, a function $d : M \times M \to \mathbb{R}^+$ such that

- Non-degeneracy: $d(x, y) = 0$ if and only if $x = y$.
- Symmetry: $d(x, y) = d(y, x)$, for every $x, y \in M$.
- Triangle inequality: $d(x, z) \le d(x, y) + d(y, z)$, for every $x, y, z \in M$.

A sequence $\{x_i\}_{i=1}^{\infty}$ of elements of a metric space (M, d) is called a *Cauchy sequence* if for every $\epsilon > 0$, there exist an integer N_ϵ, such that for every $m, n \ge N_\epsilon$, we have $d(x_m, x_n) \le \epsilon$. A metric space (M, d) is called *complete* if every Cauchy sequence has a limit in M. A metric space is *compact* if and only if every sequence in the space has a convergent subsequence.

Now that we have defined the measure spaces in Section 2, let us state the Hoölder's and Minkowski's inequalities in a more general form.

Theorem 3.1 (Hölder's inequality)**.** *Consider a measure space $\mathcal{M} = (\Omega, \mathcal{F}, \mu)$, and two reals $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$. IF the two measurable functions $f, g : \Omega \to \mathbb{C}$ are such that both $|f|^p$ and $|g|^q$ are integrable, then*

$$\left| \int f(x)\overline{g(x)} d\mu(x) \right| \le \left(\int |f(x)|^p d\mu(x) \right)^{1/p} \left(\int |g(x)|^q d\mu(x) \right)^{1/q}.$$

Theorem 3.2 (Minkowski's inequality)**.** *Consider a measure space $\mathcal{M} = (\Omega, \mathcal{F}, \mu)$, a real $p \ge 1$, and two measurable functions $f, g : \Omega \to \mathbb{C}$ such that $|f|^p$ and $|g|^p$ are both integrable. Then*

$$\left(\int |f(x) + g(x)|^p d\mu(x) \right)^{1/p} \le \left(\int |f(x)|^p d\mu(x) \right)^{1/p} + \left(\int |g(x)|^p d\mu(x) \right)^{1/p}.$$

Next we define concept of a normed space which is central to function analysis.

Definition 3.3. *A* normed space *is a pair $(V, \|\cdot\|)$, where V is a vector space over $\mathbb{R}$ or $\mathbb{C}$, and $\|\cdot\|$ is a function from V to nonnegative reals satisfying*

- **(non-degeneracy)***: $\|x\| = 0$ if and only if $x = 0$.*

- **(homogeneity)**: *For every scalar λ, and every $x \in V$, $\|\lambda x\| = |\lambda|\|x\|$.*
- **(triangle inequality)**: *For $x, y \in V$, $\|x + y\| \leq \|x\| + \|y\|$.*

We call $\|x\|$, the norm *of x. A* semi-norm *is a function similar to a norm except that it might not satisfy the non-degeneracy condition.*

The spaces $(\mathbb{C}, |\cdot|)$ and $(\mathbb{R}, |\cdot|)$ are respectively examples of 1-dimensional complex and real normed spaces.

Every normed space $(V, \|\cdot\|)$ has a metric space structure where the distance of two vectors x and y is $\|x - y\|$.

Consider two normed spaces X and Y. A *bounded operator* from X to Y, is a *linear* function $T : X \to Y$, such that

$$\|T\| := \sup_{x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X} < \infty. \tag{1}$$

The set of all bounded operators from X to Y is denoted by $B(X, Y)$. Note that the *operator norm* defined in (1) makes $B(X, Y)$ a normed space.

A *functional* on a normed space X over $\mathbb{C}$ (or $\mathbb{R}$) is a bounded linear map f from X to $\mathbb{C}$ (respectively $\mathbb{R}$), where bounded means that

$$\|f\| := \sup_{x \neq 0} \frac{|f(x)|}{\|x\|} < \infty.$$

The set of all bounded functionals on X endowed with the operator norm, is called *the dual* of X and is denoted by X^*. So for a normed space X over complex numbers, $X^* = B(X, \mathbb{C})$, and similarly for a normed space X over real numbers, $X^* = B(X, \mathbb{R})$.

For a normed space X, the set $\mathbf{B}_X := \{x : \|x\| \leq 1\}$ is called the *unit ball* of X. Note that by the triangle inequality, $\mathbf{B}_X$ is a convex set, and also by homogeneity it is symmetric around the origin, in the sense that $\|\lambda x\| = \|x\|$ for every scalar λ with $|\lambda| = 1$. The non-degeneracy condition implies that $\mathbf{B}_X$ has non-empty interior.

Every compact symmetric convex subset of $\mathbb{R}^n$ with non-empty interior is called a *convex body*. Convex bodies are in one-to-one correspondence with norms on $\mathbb{R}^n$. A convex body K corresponds to the norm $\|\cdot\|_K$ on $\mathbb{R}^n$, where

$$\|x\|_K := \sup\{\lambda \in \mathbb{R}^+ : \lambda x \in K\}.$$

Note that K is the unit ball of $\|\cdot\|_K$. For a set $K \subseteq \mathbb{R}^n$, define its *polar conjugate* as

$$K^\circ = \{x \in \mathbb{R}^n : \sum x_i y_i \leq 1, \ \forall y \in K\}. \tag{2}$$

The polar conjugate of a convex body K is a convex body, and furthermore $(K^\circ)^\circ = K$.

Consider a normed space X on $\mathbb{R}^n$. For $x \in \mathbb{R}^n$ define $T_x : \mathbb{R}^n \to \mathbb{R}$ as $T_x(y) := \sum_{i=1}^n x_i y_i$. It is easy to see that T_x is a functional on X, and furthermore every functional on X is of the form T_x for some $x \in \mathbb{R}^n$. For $x \in \mathbb{R}^n$ define $\|x\|^* := \|T_x\|$. This shows that we can identify X^* with $(\mathbb{R}^n, \|\cdot\|^*)$. Let K be the unit ball of $\|\cdot\|$. It is easy to see that K°, the polar conjugate of K, is the unit ball of $\|\cdot\|^*$.

3.1. **Hilbert Spaces.** Consider a vector space V over $\mathbb{K}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Recall that an *inner product* $\langle \cdot, \cdot \rangle$ on V, is a function from $V \times V$ to $\mathbb{K}$ that satisfies the following axioms.

- Conjugate symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
- Linearity in the first argument: $\langle ax + z, y \rangle = a\langle x, y \rangle + \langle z, y \rangle$ for $a \in \mathbb{K}$ and $x, y \in V$.
- Positive-definiteness: $\langle x, x \rangle > 0$ if and only if $x \neq 0$, and $\langle 0, 0 \rangle = 0$.

A vector space together with an inner product is called an *inner product space*.

Example 3.4. Consider a measure space $\mathcal{M} = (\Omega, \mathcal{F}, \mu)$, and let $\mathcal{H}$ be the space of measurable functions $f : \Omega \to \mathbb{C}$ such that $\int |f(x)|^2 d\mu(x) < \infty$. For two functions $f, g \in \mathcal{H}$ define

$$\langle f, g \rangle := \int f(x)\overline{g(x)} d\mu(x).$$

It is not difficult to verify that the above mentioned function is indeed an inner product. ∎

An inner product can be used to define a norm on V. For a vector $x \in V$, define $\|x\| = \sqrt{\langle x, x \rangle}$.

Lemma 3.5. *For an inner product space V, the function $\| \cdot \| : x \mapsto \sqrt{\langle x, x \rangle}$ is a norm.*

Proof. The non-degeneracy and homogeneity conditions are trivially satisfied. It remains to verify the triangle inequality. Consider two vectors $x, y \in V$ and note that by the axioms of an inner product:

$$0 \le \langle x + \lambda y, x + \lambda y \rangle = \langle x, x \rangle + |\lambda|^2 \langle y, y \rangle + \lambda \overline{\langle x, y \rangle} + \overline{\lambda} \langle x, y \rangle.$$

Now taking $\lambda := \sqrt{\frac{\langle x,x \rangle}{\langle y,y \rangle}} \times \frac{\langle x,y \rangle}{|\langle x,y \rangle|}$ will show that

$$0 \le 2\langle x, x \rangle \langle y, y \rangle - 2\sqrt{\langle x, x \rangle \langle y, y \rangle} |\langle x, y \rangle|,$$

which leads to the triangle inequality. □

A complete inner-product space is called a *Hilbert space*.

3.2. **The L_p spaces.** Consider a measure space $\mathcal{M} = (\Omega, \mathcal{F}, \mu)$. For $1 \le p < \infty$, the space $L_p(\mathcal{M})$ is the space of all functions $f : \Omega \to \mathbb{C}$ such that

$$\|f\|_p := \left(\int |f(x)|^p d\mu(x) \right)^{1/p} < \infty.$$

Strictly speaking the elements of $L_p(\mathcal{M})$ are equivalent classes. Two functions f_1 and f_2 are equivalent and are considered identical, if they agree almost everywhere or equivalently $\|f_1 - f_2\|_p = 0$.

Proposition 3.6. *For every measure space $\mathcal{M} = (\Omega, \mathcal{F}, \mu)$, $L_p(\mathcal{M})$ is a normed space.*

Proof. Non-degeneracy and homogeneity are trivial. It remains to verify the triangle inequality (or equivalently prove Minkowski's inequality). By applying Hölder's inequality:

$$\begin{aligned}
\|f + g\|_p^p &= \int |f(x) + g(x)|^p d\mu(x) = \int |f(x) + g(x)|^{p-1} |f(x) + g(x)| d\mu(x) \\
&\le \int |f(x) + g(x)|^{p-1} |f(x)| d\mu(x) + \int |f(x) + g(x)|^{p-1} |g(x)| d\mu(x) \\
&\le \left(\int |f(x) + g(x)|^p d\mu(x) \right)^{\frac{p-1}{p}} \|f\|_p + \left(\int |f(x) + g(x)|^p d\mu(x) \right)^{\frac{p-1}{p}} \|g\|_p \\
&= \|f + g\|_p^{p-1} (\|f\|_p + \|g\|_p),
\end{aligned}$$

which simplifies to the triangle inequality □

Another useful fact about the L_p norms is that when they are defined on a probability space, they are increasing.

Theorem 3.7. *Let $\mathcal{M} = (\Omega, \mathcal{F}, \mu)$ be a probability space, $1 \le p \le q \le \infty$ be real numbers, and $f \in L_q(\mathcal{M})$. Then*
$$\|f\|_p \le \|f\|_q.$$

Proof. The case $q = \infty$ is trivial. For the case $q < \infty$, by Hölder's inequality (applied with conjugate exponents $\frac{q}{p}$ and $\frac{q}{q-p}$), we have
$$\|f\|_p^p = \int |f(x)|^p \times 1 d\mu(x) \le \left(\int |f(x)|^q d\mu(x)\right)^{p/q} \left(\int 1^{\frac{q}{q-p}} d\mu(x)\right)^{\frac{q-p}{q}} = \|f\|_q^p.$$
□

Note that Theorem 3.7 does not hold when $\mathcal{M}$ is not a probability space. For example consider the set of natural numbers $\mathbb{N}$ with the counting measure. We shall use the notation $\ell_p := L_p(\mathbb{N})$. In this case the ℓ_p norms are actually decreasing.

Exercise 3.8. Let $1 \le p \le q \le \infty$. Show that for every $f \in \ell_p$, we have $\|f\|_q \le \|f\|_p$. ■

4. Basic Probabilistic Inequalities

Markov's inequality gives an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant. The application's of Markov's inequality sometimes referred to as the first moment method.

Theorem 4.1 (Markov's inequality)**.** *If X is a complex valued random variable and $a > 0$, then*
$$\Pr[|X| > a] \le \frac{\mathbb{E}[|X|]}{a}.$$

Proof. It is trivial. It follows from the definition of the expected value that
$$\mathbb{E}[|X|] \ge a\Pr[|X| > a].$$
□

In the second moment method, Chebyshev's inequality is applied to bound the probability that a random variable deviates far from the mean by its variance. Recall that the variance of a random variable is defined as
$$\mathrm{Var}[X] = \mathbb{E}\left[|X - \mathbb{E}[X]|^2\right] = \mathbb{E}[|X|^2] - |\mathbb{E}[X]|^2.$$

Theorem 4.2 (Chebyshev's inequality)**.** *If X is a complex valued random variable and $a > 0$, then*
$$\Pr[|X - \mathbb{E}[X]| > a] \le \frac{\mathrm{Var}[X]}{a^2}.$$

Proof. The theorem follows from Markov's inequality applied to the random variable $|X - \mathbb{E}[X]|^2$. □

It is possible to use Chebyshev's inequality to show that sums of independent random variables are concentrated around their expected value.

Lemma 4.3. *Let $X_1, \ldots, X_n$ be independent complex valued random variables satisfying $|X_i| \le 1$ for all $i = 1, \ldots, N$. Then*
$$\Pr\left[\left|\sum_{i=1}^n X_i - \mathbb{E}[X_i]\right| > t\right] \le \frac{n}{t^2}.$$

Proof. Denote $A = X_1 + \ldots + X_n$. Then by independence of X_i's we have

$$\mathrm{Var}[A] = \sum_{i,j=1}^{n} \mathbb{E}[X_i \overline{X_j}] - \mathbb{E}[X_i]\mathbb{E}[\overline{X_j}] = \sum_{i=1}^{n} \mathbb{E}[|X_i|^2] - |\mathbb{E}[X_i]|^2 = \sum_{i=1}^{n} \mathrm{Var}[X_i] \leq n.$$

Then Chebyshev's inequality implies the result. □

However in these situations, there are different inequalities that provide much stronger bounds compared to Chebyshev's inequality. We state two of them:

Lemma 4.4 (Chernoff Bound)**.** *Suppose that $X_1, \ldots, X_n$ are independent Bernoulli variables each occurring with probability p. Then for any $0 < t \leq np$,*

$$\Pr\left[\left|\sum_{i=1}^{n} X_i - np\right| > t\right] < 2e^{\frac{-t^2}{3np}}.$$

Lemma 4.5 (Hoeffding's Inequality)**.** *Suppose that $X_1, \ldots, X_n$ are independent random variables with $|X_i| \leq 1$ for each $1 \leq i \leq n$. Then for any $t > 0$,*

$$\Pr\left[\left|\sum_{i=1}^{n} X_i - \mathbb{E}[X_i]\right| > t\right] < 2e^{\frac{-t^2}{2n}}.$$

School of Computer Science, McGill University, Montréal, Canada
E-mail address: hatami@cs.mcgill.ca