Algorithmic Foundations of Learning **Lecture 0**

Required Knowledge

Lecturer: Patrick Rebeschini *Version: December 8, 2021*

0.1 Introduction

This is meant to give a quick overview of the type of mathematics that students are expected to know (and have good working knowledge of) in order to take this course. This also serves to introduce the notation that will be used.

Remark 0.1 (Notation) *Throughout, we use uppercase letters to denote random variables and lowercase letters to denote deterministic variables. We use cursive letters to denote sets. For a given set $\mathcal{T}$, we denote by $|\mathcal{T}|$ its cardinality. For a positive integer n, we use the notation $[n] = \{1, \dots, n\}$. For a vector $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, $p \geq 1$ we let $\|x\|_p = (\sum_{i=1}^d |x_i|^p)^{1/p}$ denote the ℓ_p norm and $\|x\|_\infty = \max_{i\in[d]} |x_i|$ denote the ℓ_∞ norm. We denote by $x^\top$ the transpose of x.*

0.2 Properties of Conditional Expectations and Probabilities

Let X, Y, Z be real-valued random variables. The following holds.

$$
\begin{aligned}
\mathbf{E}X &= \mathbf{E}\,\mathbf{E}[X|Y] && \text{(tower property or law of total expectations)}\\
\mathbf{E}[X|Y] &= \mathbf{E}[\mathbf{E}[X|Z,Y]|Y] && \text{(tower property or law of total expectations)}\\
\mathbf{E}[X|Z] &= \mathbf{E}[X|Z,Y] && \text{(if } X \text{ and } Y \text{ are independent conditionally on } Z\text{)}\\
\mathbf{E}[f(X,Y)|Y] &= \int f(x,Y)\,\mathbf{P}(X \in \mathrm{d}x|Y) && \\
\mathbf{E}[f(X,Y)|Y] &= \int f(x,Y)\,\mathbf{P}(X \in \mathrm{d}x) && \text{if } X,\, Y \text{ are independent}\\
\mathbf{E}[X|Y] &= \mathbf{E}[X] && \text{if } X,\, Y \text{ are independent}\\
\mathbf{E}[XY|Y] &= Y\,\mathbf{E}[X|Y] && \text{(``take in/out what is known'' property)}\\
\sup_{a\in\mathcal{A}} \mathbf{E}\, f(X,a) &\leq \mathbf{E} \sup_{a\in\mathcal{A}} f(X,a) && \\
\mathbf{E} \inf_{a\in\mathcal{A}} f(X,a) &\leq \inf_{a\in\mathcal{A}} \mathbf{E}\, f(X,a) &&
\end{aligned}
$$

For any event E we have $\mathbf{P}(E) = \mathbf{E}\mathbf{1}_E$, where $\mathbf{1}_E$ is the indicator function defined as $\mathbf{1}_E(w) = 1$ if $w \in E$ and $\mathbf{1}_E(w) = 0$ if $w \notin E$. The properties of conditional expectations translates into properties of conditional probabilities, as for any event E and random variable X we have $\mathbf{P}(E|X) = \mathbf{E}[\mathbf{1}_E|X]$. So, for instance, $\mathbf{P}(E) = \mathbf{E}\,\mathbf{P}(E|X)$. For any events E, F, we have $\mathbf{P}(E,F) = \mathbf{P}(E \cap F) = \mathbf{P}(E|F)\mathbf{P}(F)$. If $X_1, \dots, X_n$ are independent $\mathbf{E}[f(X_1)\cdots f(X_n)] = \mathbf{E}f(X_1)\cdots\mathbf{E}f(X_n)$.

0.3 Basic Set Operations

- For any events E and F, we have $(E \cap F)^C = E^C \cup F^C$, where E^C indicates the complementary of E.
- $\mathbf{1}_{E\cup F} \le \mathbf{1}_E + \mathbf{1}_F$.
- $\mathbf{E}\mathbf{1}_E = \mathbf{P}(E)$.
- $E \subseteq F$ if and only if $E^C \supseteq F^C$.
- Union bound: $\mathbf{P}(E \cup F) \le \mathbf{P}(E) + \mathbf{P}(F)$.
- If $E \subseteq F$, $\mathbf{P}(E) \le \mathbf{P}(F)$.
- Let $E_1, \dots, E_n$ be a countable collections of events. The following holds, with $[n] = \{1, \dots, n\}$,

$$\mathbf{P}\left(\left\{\text{There exists } i \in [n] \text{ such that } E_i \text{ is true}\right\}\right) = \mathbf{P}\left(\bigcup_{i\in[n]} E_i\right) \le \sum_{i\in[n]} \mathbf{P}(E_i).$$

$$\mathbf{P}\left(\left\{E_i \text{ is true for all } i \in [n]\right\}\right) = \mathbf{P}\left(\bigcap_{i\in[n]} E_i\right) = 1 - \mathbf{P}\left(\bigcup_{i\in[n]} E_i^C\right) \ge 1 - \sum_{i\in[n]} \mathbf{P}(E_i^C) = 1 - n + \sum_{i\in[n]} \mathbf{P}(E_i).$$

0.4 Basic Inequalities in Probability

Let X, Y be two random variables. The following holds.

- Cauchy-Schwarz's: $\mathbf{E}|XY| \le \sqrt{\mathbf{E}[X^2]}\sqrt{\mathbf{E}[Y^2]}$.
- Hölder's: $\mathbf{E}|XY| \le (\mathbf{E}[X^p])^{1/p}(\mathbf{E}[Y^q])^{1/q}$ for any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$.
- Jensen's: if f is a convex function, $\mathbf{E}f(X) \ge f(\mathbf{E}X)$. It follows that if f is concave, $\mathbf{E}f(X) \le f(\mathbf{E}X)$ (recall that if f is convex, $-f$ is concave). The same statements hold for conditional expectations: for instance, $\mathbf{E}[f(X)|Y] \ge f(\mathbf{E}[X|Y])$ if f is convex.
- $(\mathbf{E}[X^p])^{1/p} \le (\mathbf{E}[X^q])^{1/q}$ for $p \le q$.

0.5 Basic Inequalities in Linear Algebra

Let $x, y \in \mathbb{R}^d$ be d-dimensional vectors. The following holds.

- Cauchy Schwarz's: $|x^\top y| \le \|x\|_2\|y\|_2$.
- Hölder's: $|x^\top y| \le \|x\|_1\|y\|_\infty$ (In general, $|x^\top y| \le \|x\|\|y\|_*$, where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$).
- $\|x\|_\infty \le \|x\|_2 \le \|x\|_1$.
- $\|x\|_2 \le \|x\|_1 \le \sqrt{d}\|x\|_2$.
- $\|x\|_\infty \le \|x\|_1 \le d\|x\|_\infty$.
- $\|x\|_\infty \le \|x\|_2 \le \sqrt{d}\|x\|_\infty$.

0.6 Triangle Inequalities

Given a normed vector space with norm $\|\cdot\|$. For any x, y in the space, we have:

- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$.
- Reverse triangle inequality: $|\|x\| - \|y\|| \leq \|x + y\|$ (in particular, $|\|x\| - \|y\|| \leq \|x - y\|$).