MATHEMATICAL PROPERTIES OF THE STOCHASTIC APPROXIMATION AND THE MULTI-ARMED BANDIT PROBLEM

NIKOS STAMATIS

Department of Mathematics
Faculty of Sciences
Kapodistrian University of Athens

August 14, 2020

Στον μικρό Κωνσταντή
και τον ακόμα πιο μικρό Λεωνίδα.

Abstract

Despite the fact that neural networks had been used extensively for decades, a theoretical background that would explain their success was, until recently, elusive. In Chapter 2, we present the main results which settled this question, developed mostly in the early '90s. We prove Cybenko's theorem, which states that continuous and sigmoidal functions are always universal approximators, and also study some extensions of this result. Leshno et al. proved that instead of sigmoidal functions, one would suffice to use any function which is not equal to a polynomial almost everywhere. Connections with the Kolmogorov-Arnold theorem are also explored.

Chapter 3 is devoted to the study of stochastic approximation algorithms. The goal of these algorithms is to determine the fixed point of an operator when its values are not known to us, but they are revealed perturbed by some noise. They can be seen as extensions of the classical fixed point methods, like Banach's fixed point theorem. We also present the proof of the convergence of the Q-Learning algorithm which is based on this theory. The Q-Learning algorithm is a generalization of the successive approximation method, a method used extensively in the classical dynamic programming, when we have no prior information on the underlying process (transition probabilities and cost functions), but only a method to draw and observe values from it.

Lastly, in Chapter 4, we study the multi-armed bandit problem, a subfield of reinforcement learning, where the goal is to determine the most profitable action among a given set, while simultaneously, maximizing one's profit. We prove the Lai-Robbins lower bound, which shows that for a certain class of reward distributions there are limits to how fast one can reach a maximum profit, and we also present an algorithm that attains it. We conclude the chapter studying the upper confidence bound algorithm, introduced by Auer et al., which resolves several issues of the Lai-Robbins approach.

Περίληψη

Παρά το γεγονός ότι τα νευρωνικά δίκτυα χρησιμοποιούνταν επί δεκαετίες με εντυπωσιακά αποτελέσματα, η ανάπτυξη ενός θεωρητικού υπόβαθρου που θα εξηγούσε αυτήν τους την επιτυχία, είναι σχετικά πρόσφατο επίτευγμα. Στο Κεφάλαιο 2, παρουσιάζουμε τα κυριότερα αποτελέσματα που έδωσαν απάντηση σε αυτά τα ερωτήματα. Το Θεώρημα του Cybenko είναι το πρώτο σχετικό θεώρημα, σύμφωνα με το οποίο κάθε συνεχής και σιγμοειδής συνάρτηση είναι καθολικός προσεγγιστής. Οι Leshno et al., επέκτειναν το αποτέλεσμα του Cybenko, δείχνοντας ότι οποιαδήποτε μη πολυωνυμική συνάρτηση αποτελεί καθολικό προσεγγιστή. Παρουσιάζουμε επίσης μία κατασκευαστική απόδειξη στον L_2, καθώς και την προσέγγιση μέσω του Θεωρήματος των Kolmogorov και Arnold.

Το Κεφάλαιο 3 είναι αφιερωμένο στη μελέτη αλγορίθμων στοχαστικής προσέγγισης. Αυτοί οι αλγόριθμοι στοχεύουν στην εύρεση του σταθερού σημείου ενός τελεστή, όταν οι ακριβείς τιμές που παίρνει δεν είναι γνωστές σε εμάς, αλλά μας αποκαλύπτονται με την παρουσία θορύβου. Παρουσιάζουμε επίσης την απόδειξη του αλγορίθμου της Q-Μάθησης, και η οποία βασίζεται στους αλγορίθμους αυτούς. Η Q-Μάθηση αποτελεί γενίκευση μιας μεθόδου που χρησιμοποιείται ευρέως στον κλασσικό δυναμικό προγραμματισμό, της μεθόδου των διαδοχικών προσεγγίσεων, για προβλήματα στα οποία δεν έχουμε γνώση των διαφόρων παραμέτρων (πιθανότητες μετάβασης και δομή κόστους), αλλά αντίθετα μπορούμε μόνο να προσομοιώνουμε παρατηρήσεις από αυτές.

Τέλος, στο Κεφάλαιο 4, μελετάμε το πρόβλημα των multi-armed bandit, το αντικείμενο του οποίου είναι ο προσδιορισμός της πιο κερδοφόρας δράσης από ένα δοσμένο σύνολο, μαζί με την ταυτόχρονη μεγιστοποίηση του αναμενόμενου κέρδους μας. Αποδεικνύουμε το φράγμα των Lai-Robbins, σύμφωνα με το οποίο για μια συγκεκριμένη κλάση κατανομών, υπάρχουν όρια στο πόσο γρήγορα μπορούμε να πλησιάσουμε το βέλτιστο κέρδος, ενώ επίσης παρουσιάζουμε και έναν αλγόριθμο που επιτυγχάνει το φράγμα αυτό. Ο αλγόριθμος των Lai-robbins περιέχει αρκετά σκοτεινά σημεία, τα οποία προσπαθεί να απλοποιήσει η μέθοδος upper confidence bounds των Auer et al., με την οποία ολοκληρώνουμε την εργασία μας.

Acknowledgements

I would like to express my gratitude to my supervisor Professor Apostolos Burnetas for his valuable guidance during the writing of this thesis. He always had the time to discuss with me about any issue that I might had and encourage me to continue to explore paths which at first sight seemed challenging.

I am also grateful to Professors Antonis Economou and Costis Melolidakis for participating in my thesis committee. To Professor Antonis Economou for all our discussions and his support, to Yannis Dimitrakopoulos and Apostolos Burnetas for our discussions during the courses of "Dynamic Programming" and "Simulation" respectively, which largely motivated and shaped this thesis. I would also like to thank the staff members of the department for keeping a high standard in such difficult times, and especially Alkistis Ntai for her continuous help towards me and all my fellow students.

I wouldn't be writing this thesis if it wasn't for the trust of the graduate committee, Professors Antonis Economou, Costis Melolidakis, Loukia Meligotsidou and Samis Trevezas, that allowed me to participate in the graduate programme in the first place. I feel really fortunate that I was a student here and I will always think very highly of your programme and everyone involved in it.

I am indebted to my professors from my former school, Ioannis Sarantopoulos, Chrys Caroni and Andreas Boukas for having faith in me. To my friend Giorgos Katsimpas for always being there for me, for reading portions of this thesis, offering me his insights, and for all our discussions, mathematical or not, during the past years. Also to my friends Odysseas Bakas, Alkis Georgiadis - Harris and Stratoula Charitidou for all their motivation and support.

To all my friends from the graduate programme for the company (also for giving me notes everytime I was absent from class, usually without a leave), and especially Giannis Oikonomidis, Giannis Barbagiannis, Vassilis Kapetanakis and Nikos Nteits.

Special thanks to Alexandra Elbakyan. Without your sacrifices, not only this thesis, but my entire education would only be a small fraction of what it is now.

Lastly, I am grateful to my mother Amalia, my father Kostas, my brother Dimitris, his wife Mila and my nephews Konstantinos and Leonidas for all their love, patience and support.

Contents

List of Figures

List of Tables

Notation

A_f	The set of discontinuity points of the function f.
B_X	The closed unit ball of a normed space $(X, \|\cdot\|)$.
$C(X)$	The space of continuous functions on X.
$C^{(k)}(X)$	The space of differentiable functions on X for which the first k derivatives are continuous.
$C^\infty(X)$	The space of smooth functions on X.
$C_C^\infty(X)$	The space of smooth functions with compact support on X.
$d(A)$	The natural density of a subset A of $\mathbb{N}$.
$\mathbb{E}[X \mid \mathcal{F}]$	The conditional expectation of a random variable X with respect to a sigma-algebra $\mathcal{F}$.
I_n	The n-th dimensional cube $I_n = [0,1]^n$ in $\mathbb{R}^n$.
$L_p(\Omega)$	The space of p-integrable functions, $1 \le p < \infty$, on Ω.
$L_\infty(\Omega)$	The space of essentially bounded functions on Ω.
$L_{\text{loc}}^\infty(\Omega)$	The space of functions which belong to $L_\infty(K)$ for every compact K subset of Ω.
ℓ_p	The space of p-summable real sequences for $1 \le p < \infty$.
ℓ_∞	The space of bounded real sequences.
$P_F(x)$	The projection of the point x to the set F, most likely in a Hilbert space.
$\left(\mathbb{R}^n, \|\cdot\|_\xi\right)$	The weighted supremum norm $\|x\|_\xi = \sup_{i=1,\dots,n} \frac{\|x_i\|}{\xi_i}$ on $\mathbb{R}^n$, induced by the strictly positive vector ξ.
$\text{orb}(x, T)$	The orbit $\left\{T^n x : n \in \mathbb{N}\right\}$ of a point x under the operator T defined on some vector space, or some metric space, depending on the context.
S_n	The n-th partial sum, $S_n = \sum_{k=1}^n a_k$, of a sequence $(a_k)_k$.
$\Sigma_n(\sigma)$	The set of all possible functions $s : \mathbb{R}^n \to \mathbb{R}$ which can be generated by a neural network having σ as an activation function.
(X, ρ)	A metric space.
$(X, \|\cdot\|)$	A normed space.
X^*	The topological dual of a normed space $(X, \|\cdot\|)$.
(X, τ)	A topological space, or a topological vector space, depending on the context.

1

Εκτενής Περίληψη

1.1 *Θεωρήματα καθολικής προσέγγισης*

Τα νευρωνικά δίκτυα έχουν μια πλούσια ιστορία η οποία ξεκίνησε τη δεκαετία του '40, με την πρώτη τους θεωρητική σύλληψη. Σταδιακά, και με την παράλληλη εξέλιξη των υπολογιστικών δυνατοτήτων, η χρήση νευρωνικών δικτύων έδωσε εντυπωσιακά αποτελέσματα στην επίλυση ιδιαίτερα πολύπλοκων προβλημάτων σε ευρύ φάσμα εφαρμογών. Όμως, ενώ η αποτελεσματικότητα των νευρωνικών δικτύων είχε επιβεβαιωθεί απο την πρακτική εμπειρία, η θεωρητική αιτιολόγησή της είναι ένα σχετικά σύγχρονο επίτευγμα το οποίο ανέδειξε τη σύνδεση της θεωρίας των νευρωνικών δικτύων με κλάδους των θεωρητικών μαθηματικών.

Για να οριστεί ένα νευρωνικό δίκτυο, χρειαζόμαστε δύο κύρια δομικά συστατικά, τις σιγμοειδείς συναρτήσεις και τους αφφινικούς μετασχηματισμούς.

Ορισμός 1.1.1: Μια συνάρτηση $\sigma : \mathbb{R} \to \mathbb{R}$ καλείται *σιγμοειδής*, εαν

$$\sigma(t) \longrightarrow \begin{cases} 0, & \text{για } t \to -\infty, \\ 1, & \text{για } t \to +\infty. \end{cases} \tag{1.1.1}$$

Ορισμός 1.1.2: Μια συνάρτηση $A : X \to Y$ μεταξύ δύο διανυσματικών χώρων ονομάζεται *αφφινική*, εαν ισχύει ότι $A\left(\sum_{i=1}^{n} \lambda_i x_i\right) = \sum_{i=1}^{n} \lambda_i A(x_i)$ για κάθε $n \in \mathbb{N}$, $x_i \in X$ και κάθε $\lambda_i \in \mathbb{R}$ με $\sum_{i=1}^{n} \lambda_i = 1$.

Ως νευρωνικό δίκτυο, θεωρούμε κάθε συνάρτηση s της μορφής $s(x) = Tx$, όπου

$$T = A_{m+1} S_m A_m \cdots A_2 S_1 A_1 \tag{1.1.2}$$

είναι ένας τελεστής ο οποίος ορίζεται απο τις διαδοχικές συνθέσεις αφφινικών μετασχηματισμών $A_i : \mathbb{R}^{d_{i-1}} \to \mathbb{R}^{d_i}$ με σιγμοειδείς συναρτήσεις S_i. Ο αριθμός m μετρά το πλήθος των στρωμάτων του δικτύου, ενώ οι αριθμοί d_i το πλήθος των κόμβων που εμφανίζονται σε κάθε στρώμα. Θα συμβολίζουμε με

$$\Sigma_n(\sigma) = \Big\{ s : \mathbb{R}^n \to \mathbb{R} : s(x) = \sum_{j=1}^{N} a_j \sigma(w_j^{\intercal} x + \theta_j) : \tag{1.1.3}$$

$$N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \Big\}, \tag{1.1.4}$$

το σύνολο των συναρτήσεων s που δύνανται να αναπαρασταθούν από ένα νευρωνικό δίκτυο βάθους ένα, το οποίο έχει ως συνάρτηση ενεργοποίησης τη σ.

Γενικά, δεν είναι σωστό ότι κάθε συνεχής συνάρτηση μπορεί να γραφεί στη μορφή (1.1.2), και άλλωστε δε θα περιμέναμε εκ των προτέρων να ίσχυε κάτι τέτοιο. Όμως η πολυετής αποτελεσματική χρήση των νευρωνικών δικτύων σε πραγματικά προβλήματα, αποτέλεσε ισχυρή ένδειξη ότι κάθε συνεχής συνάρτηση θα μπορούσε να προσεγγισθεί από τέτοιες συναρτήσεις.

1.1.1 *Το Θεώρημα του Cybenko*

Θεωρούμε ότι οι εμπλεκόμενες συναρτήσεις ορίζονται στο $I_n = [0,1]^n$, αντί για ολόκληρο το $\mathbb{R}^n$, ούτως ώστε να εκμεταλλευτούμε τις ιδιότητες του χώρου $C(I_n)$. Από μαθηματική σκοπιά, το ερώτημα της καθολικής προσέγγισης διατυπώνεται ως εξής: *Για ποιες συναρτήσεις ενεργοποίησης σ ισχύει ότι το σύνολο*

$$\begin{aligned}\Sigma_n(\sigma) = \Big\{ f : I_n \to \mathbb{R} : f(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j) : \\ N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \Big\} \\ = \operatorname{span}\{ f : f(x) = \sigma\left(w^\intercal x + \theta\right) \ \ \textit{για } w \in \mathbb{R}^n, \theta \in \mathbb{R} \}.\end{aligned}$$

είναι πυκνό στον $C(I_n)$; Συναρτήσεις με την ιδιότητα αυτή, καλούνται *καθολικοί προσεγγιστές*. Παραδείγματα συγκερκιμένων καθολικών προσεγγιστών ήταν ήδη γνωστά από τα μέσα της δεκαετίας του '80, όμως το πρώτο πραγματικά γενικό αποτέλεσμα ανήκε στον George Cybenko, ο οποίος το 1989 [Cyb89] απέδειξε ότι κάθε συνεχής σιγμοειδής συνάρτηση είναι καθολικός προσεγγιστής.

Θεώρημα 1.1.1 (Cybenko): *Για κάθε συνεχή σιγμοειδή συνάρτηση σ, το σύνολο $\Sigma_n(\sigma)$ είναι πυκνό στον $C(I_n)$.*

Η απόδειξη του θεωρήματος είναι υπαρξιακή, και ουσιαστικά αποτελεί μια σχετικά απλή, αλλά ιδιαίτερα εντυπωσιακή, εφαρμογή του θεωρήματος Hahn - Banach. Το σύνολο $Y = \Sigma_n(\sigma)$ αποτελεί γραμμικό υπόχωρο του $C(I_n)$. Εαν δεν ήταν πυκνός, τότε από το Θεώρημα Hahn-Banach, θα υπήρχε κάποιο γραμμικό και φραγμένο συναρτησιακό $0 \neq x^* \in C(I_n)^*$ με $x^*(Y) = 0$.

Από το Θεώρημα Αναπαράστασης του Riesz, ο δυϊκός του $C(I_n)$ μπορεί να ταυτιστεί με το χώρο πεπερασμένων, προσημασμένων μέτρων Borel του I_n, επομένως θα έπρεπε να υπάρχει ένα τέτοιο μη μηδενικό μέτρο μ, με την ιδιότητα ότι

$$\int_{I_n} f(x) d\mu(x) = 0 \tag{1.1.5}$$

για κάθε $f \in Y$. Η απόδειξη ολοκληρώνεται καταλήγωντας σε άτοπο, και βασίζεται σε ένα επιχείρημα αρμονικής ανάλυσης που παρουσιάζουμε αναλυτικά (Theorem 2.2.6).

1.1.1.1 *Προβλήματα κατηγοριοποίησης*

Ίσως μια εξίσου σημαντική κλάση προβλημάτων, είναι τα προβλήματα κατηγοριοποίησης. Σε αυτά τα προβλήματα υποθέτουμε ότι υπάρχουν k το πλήθος διαφορετικοί πληθυσμοί από τους οποίους λαμβάνουμε παρατηρήσεις, και ο σκοπός μας είναι να αποφανθούμε από ποιον πληθυσμό προήθλε η κάθε μία. Από μαθηματική άποψη, μας ενδιαφέρει η εκμάθηση μιας συνάρτησης $f : I_n \to \{1, \dots, k\}$, η οποία σε κάθε σημείο του μοναδιαίου υπερκύβου αναθέτει την αντίστοιχη κατηγορία από την οποία έχει προέλθει.

Μιας και οι συγκεκριμένες συναρτήσεις είναι πάντοτε ασυνεχείς (δες σελ. 26), τα προβλήματα κατηγοριοποίησης χρειάζονται μια μικρή τροποποίηση ώστε να μπορέσουν να ενταχθούν στο προηγούμενο πλαίσιο. Στην Παράγραφο 2.2.1 εξηγούμε πώς το θεώρημα του Cybenko δύναται να εφαρμοστεί σε αυτή την κατηγορία προβλημάτων. Η βασική ιδέα είναι ότι αντί για την ασυνεχή f, μπορούμε να προσεγγίσουμε μια συνεχή συνάρτηση που είναι αρκετά κοντά της, η ύπαρξη της οποίας εξασφαλίζεται από το Θεώρημα του Lusin. Το τίμημα που πληρώνουμε, είναι ότι υπάρχει πάντα μια αυθαίρετα μικρή, αλλά θετική πιθανότητα να κάνουμε λάθος κατά την κατηγοριοποίηση.

1.1.2 *Μη-πολυωνυμικές συναρτήσεις ενεργοποίησης*

Όπως έδειξε ο Cybenko, κάθε *συνεχής* και *σιγμοειδής* συνάρτηση είναι καθολικός προσεγγιστής. Ένα φυσιολογικό ερώτημα είναι το κατά πόσον μπορούν να χαλαρώσουν αυτές οι δύο υποθέσεις και η συνάρτηση ενεργοποίησης να εξακολουθήσει να έχει αυτήν την ιδιότητα.

Στην Παράγραφο 2.3 προυσιάζουμε αναλυτικά ένα αποτέλεσμα των Leshno, Lin, Pinkus και Schocken [LLPS93], οι οποίοι έδωσαν έναν εντυπωσιακό χαρακτηρισμό για τις συναρτήσεις ενεργοποίησης οι οποίες αποτελούν καθολικούς προσεγγιστές. Απέδειξαν ότι οποιαδήποτε συνάρτηση είναι καθολικός προσεγγιστής, αρκεί να μην είναι ίση με κάποιο πολυώνυμο σχεδόν παντού.

Η μία κατεύθυνση του θεωρήματος είναι αρκετά προφανής. Αν η συνάρτηση ενεργοποίησης σ είναι πολυώνυμο βαθμού k, τότε ο υπόχωρος $\Sigma_n(\sigma)$ θα αποτελείται με τη σειρά του από πολυώνυμα βαθμού το πολύ k, επομένως αποκλείεται να είναι πυκνός στις συνεχείς συναρτήσεις. Η αντίστροφη κατεύθυνση είναι ιδιαίτερα επίπονη και χρησιμοποιεί ισχυρά εργαλεία από την συναρτησιακή και αρμονική ανάλυση.

Το αποτέλεσμα των Leshno et al. αναδεικνύει ότι η ουσία ενός νευρωνικού δικτύου, τουλάχιστον όσον αφορά στις προσεγγιστικές του δυνατότητες, δεν είναι συνυφασμένη με τις σιγμοειδείς συναρτήσεις. Οποιαδήποτε (μη πολυωνυμική) συνάρτηση μπορεί να χρησιμοποιηθεί στη θέση μιας σιγμοειδούς. Αντίθετα, οι προσεγγιστικές δυνατότες των νευρωνικών δικτύων πηγάζουν από τη δομή της σύνθεσης ανάμεσα στα διάφορα επίπεδα του δικτύου.

1.1.3 *Κατασκευαστικές αποδείξεις*

Η απόδειξη του Cybenko είναι καθαρά υπαρξιακή, γεγονός το οποίο, ιδιαίτερα αναλογιστεί κανείς την εφαρμοσμένη φύση του όλου προβλήματος, όχι μόνο δεν έληξε το ζήτημα της καθολικής προσέγγισης, αλλά ώθησε στην αναζήτηση κατασκευαστικών αποδείξεων και επεκτάσεων. Έχοντας πια ως δεδομένο ότι η καθολική προσέγγιση ήταν εφικτή, η εύρεση κατασκευαστικών μεθόδων όχι απλά φάνταζε ρεαλιστική, αλλά και θα συνέβαλε σε μια καλύτερη κατανόηση του πώς ακριβώς επιτυγχάνεται η καθολική προσέγγιση, ένα ζήτημα στο οποίο η απόδειξη του Cybenko δεν ήταν σε θέση να ρίξει φως.

Η πρώτη κατασκευαστική απόδειξη δόθηκε το 1992 από τους Chen, Chen και Liu [CCL91], οι οποίοι μάλιστα χαλάρωσαν την υπόθεση περί συνέχειας της συνάρτησης ενεργοποίησης. Ίδιας φιλοσοφίας μπορεί να θεωρηθεί και η προσέγγιση μέσω του Θεωρήματος Υπέρθεσης των Kolmogorov και Arnold που περιγράφουμε στην επόμενη παράγραφο και αναπτύχθηκε παράλληλα με το άρθρο των Chen et al.

Στην Παράγραφο 2.4, παρουσιάζουμε μια κατασκευαστική προσέγγιση στον $L_2(X)$, η οποία αναπτύχθηκε από τους Kwok and Yeung [KY97]. Υποθέτουμε την ύπαρξη κάποιου συνόλου Γ τέτοιο ώστε $\overline{\operatorname{span}\Gamma} = L_2(X)$. Για παράδειγμα, το Γ θα μπορούσε να περιέχει τις συναρτήσεις της μορφής

$$\left\{f : \mathbb{R}^n \to \mathbb{R} : f(x) = \sigma\left(w^\intercal x + \theta\right) \text{ for } w \in \mathbb{R}^n, \theta \in \mathbb{R}\right\}$$

για μη-πολυνωμιακή συνάρτηση σ, οι οποίες ήταν ήδη γνωστό ότι έχουν αυτή την ιδιότητα. Η διατύπωση του θεωρήματος είναι αρκετά γενική, ώστε να μη χρειάζεται να ασχοληθούμε με το τι είδους στοιχεία περιέχονται στο Γ, αλλά είναι σίγουρα χρήσιμο να θυμόμαστε ότι το σύνολο αυτό παίζει το ρόλο του νευρωνικού δικτύου.

Θεώρημα 1.1.2 (Kwok-Yeung): *Έστω $X \subseteq \mathbb{R}^d$ συμπαγές και $\Gamma \subseteq L_2(X)$ σύνολο με την ιδιότητα ότι ο υπόχωρος που παράγει είναι πυκνός στον $L_2(X)$. Έστω επίσης $f \in L_2(X)$. Μπορούμε να κατασκευάσουμε ακολουθία $(f_n)_n$ στο $\operatorname{span}\Gamma$, τέτοια ώστε $f_n \to f$.*

Η βασική ιδέα για την επαγωγική κατασκευή της ακολουθίας $(f_n)_n$ είναι η ακόλουθη: Στο πρώτο βήμα, επιλέγουμε αυθαίρετα μια f_1 η οποία ανήκει στο Γ. Έστω ότι μετά το n-στό βήμα έχουμε αποφασίσει ότι οι συναρτήσεις $g_1, \dots, g_n \in \Gamma$ πρέπει να παράγουν την f_n, δηλαδή ότι $f_n \in \operatorname{span}\{g_1, \dots, g_n\} =: F_n$. Για να αποφασίσουμε τους συντελεστές β_i που θα εμφανίζονται στο γραμμικό συνδυασμό $f_n = \sum_{i=1}^n \beta_i g_i$, παίρνουμε την f_n να είναι ίση με την προβολή της f στον υπόχωρο F_n.[1]

Συγκεκριμένα, σύμφωνα με την αντίστοιχη θεωρία, υπολογίζονται οι συντελεστές της $f_n = \sum_{i=1}^n \lambda_i e_i$ ως προς κάποια ορθομοναδιάια βάση $(e_i)_{i=1}^n$ του F_n

[1] Αυτό φυσικά προϋποθέτει ότι γνωρίζουμε ποια είναι η f.

(Πρόταση A.3.9), και στη συνέχεια η f_n ξαναγράφεται ως προς τα g_i's. Παίρνοντας την προβολή της f στον υπόχωρο F_n, ουσιαστικά εξασφαλίζουμε ότι η f_n είναι το στοιχείο του F_n το οποίο είναι πλησιέστερο στη συνάρτηση f την οποία επιθυμούμε να προσεγγίσουμε.

Για να αποφασίσουμε το επόμενο στοιχείο g_{n+1} που θα προστεθεί στο σύνολο $\{g_1, \ldots, g_n\}$, επιλέγουμε το $g^* = g_{n+1} \in \Gamma$ που ελαχιστοποιεί το σφάλμα

$$\|f - (f_{n-1} + \beta g)\|$$

για $\beta \in \mathbb{R}$ και $g \in \Gamma$, και συνεχίζουμε με τον ίδιο τρόπο. Σε κάθε επανάληψη, το σύνολο $\{g_1, \ldots, g_n\}$ εμπλουτίζεται, και οι συντελεστές των g_i επαναϋπολογίζονται, ώστε να ελαχιστοποιήσουν το σφάλμα.

Αξίζει να σημειωθεί ότι υπάρχουν πολλές παραλλαγές αυτού του αλγορίθμου. Μάλιστα, σε κάποιους από αυτούς, το πρόβλημα της ελαχιστοποίησης που πρέπει να επιλυθεί κατά την επιλογή του καινούργιου στοιχείου g_{n+1}, προσπερνάται με έναν αρκετά ευφάνταστο τρόπο: Το καινούργιο στοιχείο επιλέγεται τυχαιοποιώντας. Κάτω από ήπιες υποθέσεις, η ακολουθία που κατασκευάζεται συγκλίνει στη συνάρτηση που επιθυμούμε με πιθανότητα ένα.

1.1.4 *Το Θεώρημα των Kolmogorov - Arnold*

Στην προσπάθειά τους να επιλύσουν το 13ο πρόβλημα του Hilbert, οι Andrey Kolmogorov και Vladimir Arnold, δημοσίευσαν στα τέλη της δεκαετίας του '50 μία σειρά από άρθρα, στα οποία όχι μόνο έδωσαν απάντηση στο εν λόγω πρόβλημα, αλλά και οδήγησαν στην ανακάλυψη ενός εκ των σπουδαιότερων αποτελεσμάτων των σύγρονων μαθηματικών.

Ο Hilbert είχε εικάσει ότι οι ρίζες της εξίσωσης $x^7 + ax^3 + bx^2 + cx + 1 = 0$, ιδωμένες σαν συνάρτηση των τριών μεταβλητών a, b, c, δεν μπορούσαν να γραφούν ως σύνθεση συναρτήσεων *δύο* μεταβλητών. Ο Arnold [Arn57] πρώτος κατέρριψε την εικασία, αλλά ο Kolmogorov [Kol57] προχώρησε ένα βήμα παραπέρα, αποδεικνύοντας ότι κάθε συνάρτηση d μεταβλητών, ορισμένη στον μοναδιαίο υπερκύβο $f : [0,1]^d \to \mathbb{R}$, μπορεί να γραφεί ως υπέρθεση συναρτήσεων μόλις μίας μεταβλητής.

Θεώρημα 1.1.3 (Kolmogorov-Arnold): *Υπάρχουν σταθερές $\lambda_1, \ldots, \lambda_d \in \mathbb{R}$ για τις οποίες $\sum_{j=1}^{d} \lambda_j \leq 1$, και συνεχείς συναρτήσεις $\phi_1, \ldots, \phi_{2d+1}$ από το $[0,1]$ στον εαυτό του, με την ιδιότητα ότι κάθε $f \in C[0,1]^d$ μπορεί να γραφτεί ως*

$$f(x_1, \ldots, x_d) = \sum_{i=1}^{2d+1} g\left(\sum_{j=1}^{d} \lambda_j \phi_i(x_j)\right), \tag{1.1.6}$$

όπου $g \in C[0,1]$ μια συνάρτηση που εξαρτάται από την f.

Το 1987, δύο χρόνια προτού ο George Cybenko αποδείξει το πρώτο θεώρημα καθολικής προσέγγισης, ο Robert Hecht-Nielsen παρατήρησε σε ένα σημείωμά

του [Hec87] την ομοιότητα μεταξύ της έκφρασης (1.1.6) και των συναρτήσεων που απαρτίζουν ένα νευρωνικό δίκτυο

$$s(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j). \tag{1.1.7}$$

Με μεγάλο ενθουσιασμό, πρότεινε το θεώρημα υπέρθεσης των Kolmogorov και Arnold ως ένα πιθανό εργαλείο για την υλοποίηση του στόχου της καθολικής προσέγγισης, σημειώνοντας παράλληλα ότι, εαν αυτό ήταν όντως εφικτό, τότε θα αποτελούσε και την πρώτη εφαρμογή του θεωρήματος υπέρθεσης, τριάντα ολόκληρα χρόνια μετά την ανακάλυψή του.

Η ανταπόκριση στο σημείωμα του Nielsen ήταν ιδιαίτερα θερμή. Πρώτα οι Girosi και Poggio (1989), σε ένα άρθρο τους με τίτλο "*Kolmogorov's theorem is irrelevant*" [GP89], έδειξαν ότι είναι αδύνατο να επιτύχει κανείς την αναπαράσταση (1.1.6) χρησιμοποιώντας νευρωνικά δίκτυα, επειδή οι εμπλεκόμενες συναρτήσεις του θεωρήματος υπέρθεσης είναι παθολογικές, εν αντιθέσει με τις καλώς συμπεριφερόμενες σιγμοειδείς συναρτήσεις που χρησιμοποιεί ένα νευρωνικό δίκτυο.

Όμως, παρά την πρώτη αρνητική απάντηση, το ποτήρι αποδείχθηκε μισογεμάτο. Μπορεί το θεώρημα υπέρθεσης να μην ενέπιπτε στα πλαίσια των νευρωνικών δικτύων όσον αφορά στην αναπαράσταση συναρτήσεων, εν τούτοις το εξίσου σημαντικό ζήτημα της προσέγγισης συναρτήσεων παραμένε ανοικτό και απαντήθηκε καταφατικά δύο χρόνια αργότερα από την Věra Kůrková ([Kur91], [Kur92]), στο άρθρο της με τον εξίσου εύγλωττο τίτλο "*Kolmogorov's theorem is relevant*". Η ιδέα της ήταν να χρησιμοποιήσει σιγμοειδείς συναρτήσεις για να προσεγγίσει σημειακά όλες τις συναρτήσεις που εμπλέκονται στο θεώρημα υπέρθεσης και στη συνέχεια, επικαλώντας το, να συμπεράνει την καθολική προσέγγιση των νευρωνικών δικτύων.

Τελικά, όχι απλά το θεώρημα υπέρθεσης αποδείχθηκε σχετικό, αλλά επιπρόσθετα η ίδια η ιδιότητα της καθολικής προσέγγισης των νευρωνικών δικτύων μπορούσε να ιδωθεί ως μια ιδιαίτερη έκφανσή του. Η δουλειά της Kůrková αποτέλεσε την αφετηρία για μια σειρά αποτελεσμάτων προς διάφορες κατευθύνσεις, στα οποία το θεώρημα υπέρθεσης διαδραμάτιζε κεντρικό ρόλο. Στην εργασία μας, παρουσιάζουμε ένα άρθρο των Vitaly Maiorov και Allan Pinkus (1999) [MP99] στο οποίο εξετάζουν το πρόβλημα της καθολικής προσέγγισης από μια διαφορετική σκοπιά.

Εως τώρα τα ερωτήματα που έχουμε διατυπώσει, ασχολούνται με το ποιες συναρτήσεις ενεργοποίησης μπορούν να χρησιμοποιηθούν ως καθολικοί προσεγγιστές. Οι Maiorov και Pinkus, από την άλλη, εξετάζουν το κατά πόσον υπάρχει κάποια συνάρτηση ενεργοποίησης για την οποία να επιτυγχάνεται πάντα καθολική προσέγγιση με σχετικά λίγους κόμβους. Πράγματι, με τη βοήθεια του θεωρήματος υπέρθεσης, βρίσκουν μια αρκετά παθολογική συνάρτηση ενεργοποίησης, η οποία μπορεί να προσεγγίσει οποιαδήποτε συνεχή συνάρτηση χρησιμοποιώντας μονάχα δύο στρώματα και σχετικά λίγους κόμβους, ο αριθμός των οποίων εξαρτάται από τη διάσταση d του προβλήματος:

Θεώρημα 1.1.4 (Maiorov-Pinkus): [MP99] *Υπάρχει λεία, σιγμοειδής συνάρτηση ενεργοποίησης σ, τέτοια ώστε για κάθε $d \in \mathbb{N}$, κάθε συμπαγές $K \subseteq \mathbb{R}^d$, κάθε $f \in C(K)$ και $\varepsilon > 0$, να υπάρχουν πραγματικές σταθερές d_i, c_{ij}, θ_{ij}, γ_i και διανύσματα $w_{ij} \in \mathbb{R}^d$, με*

$$\left| f(x) - \sum_{i=1}^{6d+3} d_i \sigma \left(\sum_{j=1}^{3d} c_{ij} \sigma \left(w_{ij}^{\mathsf{T}} x + \theta_{ij} \right) + \gamma_i \right) \right| < \varepsilon \tag{1.1.8}$$

για κάθε $x \in K$.

Απόδειξη: Δες Theorem 2.5.2. ■

1.2 Στοχαστική προσέγγιση

Ένας συνηθισμένος τρόπος επίλυσης δύσκολων προβλημάτων βελτιστοποίησης, είναι μέσω της χρήσης επαναληπτικών προσεγγιστικών μεθόδων. Η στοχαστική προσέγγιση αποτελεί επέκταση των μεθόδων αυτών, όταν στο πρόβλημα υπάρχει κάποιου είδους τυχαιότητα.

1.2.1 Ο αλγόριθμος Robbins-Monro

Στην Παράγραφο 3.1, περιγράφουμε τον αλγόριθμο Robbins-Monro [RM51], ο οποίος ιστορικά αποτελεί και το πρώτο αποτέλεσμα στοχαστική προσέγγισης. Το πλαίσιο στο οποίο δουλεύουμε είναι το εξής: Υποθέτουμε ότι για κάθε $x \in \mathbb{R}$, παρτηρούμε μια τυχαία μεταβλητή $Y = Y(x)$ με κατανομή $P[Y(x) \leq y] = H(y \mid x)$ και αναμενόμενη τιμή $M(x) = \mathbb{E}[Y \mid X = x] = \int_{\mathbb{R}} y dH(y \mid x)$. Εν γένει, η ακριβής μορφή της $M(x)$, ή ακόμα και της κατανομής $H(y \mid x)$ δεν είναι γνωστές, αλλά υποθέτουμε ότι μπορούμε να προσομοιώνουμε από την εν λόγω κατανομή για οποιαδήποτε τιμή του x. Το ζητούμενο είναι να βρεθεί μια μέθοδος επίλυσης εξισώσεων της μορφής $M(\theta) = a$ ως προς θ. Δηλαδή, μας ενδιαφέρει να βρούμε ένα κατώφλι, πέρα από το οποίο η απόκριση της τυχαίας μεταβλητής θα είναι τουλάχιστον a.

Στον αλγόριθμο Robbins-Monro, κατασκευάζεται αναδρομικά μια ακολουθία $(x_n)_n$ σύμφωνα με τον τύπο

$$x_{n+1} = x_n + a_n(a - y_n), \tag{1.2.1}$$

όπου y_n είναι μια παρατήρηση που προσομοιώθηκε από την κατανομή $H(y \mid x_n)$ και $(a_n)_n$ μια προκαθορισμένη ακολουθία μη αρνητικών αριθμών. Υπο συγκεκριμένες υποθέσεις για τις κατανομές $Y(x)$ και την ακολουθία $(a_n)_n$, εξασφαλίζεται η σύγκλιση της $(x_n)_n$ στο θ κατά πιθανότητα.

Θεώρημα 1.2.1 (Robbins-Monro): *Υποθέτουμε ότι υπάρχει σταθερά $C > 0$ με $P[|Y(x)| \leq C] = 1$ για κάθε $x \in \mathbb{R}$. Επιπλέον, η συνάρτηση $M(x) = \mathbb{E}[Y \mid X = x]$*

είναι αύξουσα, με $M(\theta) = a$ *και* $M'(\theta) > 0$. *Αν η μη-αρνητική ακολουθία* $(a_n)_n$ *ανήκει στον* $\ell_2 \setminus \ell_1$, *τότε η ακολουθία* $(x_n)_n$ *του αλγορίθμου Robbins-Monro*

$$x_{n+1} = x_n + a_n(a - y_n), \tag{1.2.2}$$

συγκλίνει στο θ *κατά πιθανότητα,* $x_n \xrightarrow{P} \theta$.

Η ποσότητα x_n στη σχέση (1.2.2) αντιπροσωπεύει την τρέχουσα εκτίμηση για το θ. Αυτή η τιμή διορθώνεται κατά τον παράγοντα $a_n(a - y_n)$, έτσι ώστε να δώσει την καινούργια εκτίμηση x_{n+1}. Για την ακρίβεια, η τιμή $x_{n+1} = (1 - a_n)x_n + a_n(a - y_n + x_n)$ αποτελεί κυρτό συνδυασμό της προηγούμενης εκτίμησης x_n και της προτεινόμενης διόρθωσης $a - y_n + x_n$. Οι δύο αυτές τιμές, x_n και $a - y_n + x_n$, αποτελούν τις δύο ακραίες προτάσεις για την x_{n+1}. Η πρώτη υποδεικνύει ότι η x_{n+1} πρέπει να αγνοήσει πλήρως τον διορθωτικό όρο $a - y_n$, ενώ η δεύτερη ότι θα πρέπει να τον αποδεχθεί εξ' ολοκλήρου. Η ύπαρξη της σταθεράς a_n καθορίζει το βάρος που θα δοθεί στε κάθε μία από τις δύο ακραίες αυτές προτάσεις.

Ο ρόλος της ακολουθίας $(a_n)_n$ είναι ιδιαίτερος, καθώς επιμερίζεται μεταξύ δύο αντικρουόμενων αλλά επιθυμητών συμπεριφορών. Από τη μία πλευρά, οι όροι της $(a_n)_n$ θα πρέπει να είναι αρκετά μεγάλοι ώστε να ληφθούν υπόψιν οι διορθώσεις $a - y_n$, αλλά και σχετικά μικροί, ώστε να μη χαραμιστεί η πρόοδος του αλγορίθμου και η οποία εκφράζεται από την τρέχουσα τιμή του x_n. Αυτή είναι μία έκφανση του διλήμματος εξερεύνησης - εκμετάλλευσης το οποίο θα ξανασυναντήσουμε στο κεφάλιο των multi-armed bandit.

Οι Robbins και Monro, επιτυγχάνουν αυτόν τον συμβιβασμό, επιλέγοντας την $(a_n)_n$ να είναι τετραγωνικά αθροίσιμη, αλλά όχι απολύτως αθροίσιμη, δηλαδή δουλεύοντας με ακολουθίες οι οποίες είναι μεν αρκετά μικρές ώστε να συγκλίνουν στο μηδέν, αλλά με σχετικά αργό ρυθμό.

1.2.2 *Εύρεση σταθερών σημείων υπό τυχαιότητα*

Πολλά απαιτητικά προβλήματα στα μαθηματικά, ιδιαίτερα ζητήματα ύπαρξης, ανάγονται στην εύρεση σταθερού σημείου για κάποια κατάλληλη συνάρτηση. Στην Παράγραφο 3.2, παρουσιάζουμε ένα εργαλείο που μας επιτρέπει να βρίσκουμε το σταθερό σημείο μιας συνάρτησης υπό την παρουσία τυχαιότητας.

Θεωρούμε έναν (όχι απαραίτητα γραμμικό) τελεστή $H : X \to X$ που δρα πάνω στο διανυσματικό χώρο X και ο οποίος γνωρίζουμε ότι έχει μοναδικό σταθερό σημείο. Μια συνήθης μέθοδος για την προσέγγιση του σημείου αυτού, θα ήταν να χρησιμοποιήσουμε την αναδρομική ακολουθία $(x_n)_n$ που ορίζεται από τη σχέση $x_{n+1} = Hx_n$. Η ακολουθία αυτή ξεκινά από κάποιο αυθαίρετο $x_1 \in X$, και εν συνεχεία ακολουθεί την τροχιά αυτού του σημείου μέσω του τελεστή H. Υπό κατάλληλες προϋποθέσεις, για παράδειγμα όταν ο H είναι συστολή και ο X χώρος Banach, η ακολουθία $(x_n)_n$ όντως συγκλίνει στο ζητούμενο σταθερό σημείο του H.

Στο πρόβλημα που μελετάμε όμως, λόγω της παρουσίας θορύβου, είναι αδύνατο να γνωρίζουμε την ακριβή τιμή Hx_n κάθε τέτοιας επανάληψης. Αντίθετα,

παρατηρούμε μια τιμή $Hx_n + w_n$, όπου ο w_n είναι μια τυχαία μεταβλητή που παίζει το ρόλο του θορύβου. Η ύπαρξη αυτού του όρου, μας αναγκάζει να προσαρμόσουμε τον τρόπο που ορίζουμε την αναδρομική ακολουθία μας. Πιο συγκεκριμένα, και εναρμονισμένοι με τη προσέγγιση των Robbins-Monro, ορίζουμε την $(x_n)_n$ σύμφωνα με τη σχέση

$$x_{n+1} = (1-\gamma_n)x_n + \gamma_n(Hx_n + w_n) \tag{1.2.3}$$

όπου $(\gamma_n)_n$ είναι κατάλληλη ακολουθία στο $(0,1]$. Η ακολουθία αυτή, δεν είναι απαραίτητο να είναι προκαθορισμένη, αλλά μπορεί ο κάθε όρος της να εξαρτάται από την ιστορία της ανέλιξης μέχρι εκείνη τη στιγμή.

Τα δύο κύρια αποτελέσματα που αποδεικνύουμε, αφορούν σε δύο μεγάλες κλάσεις τελεστών $H : \mathbb{R}^N \to \mathbb{R}^N$, τις ψευδοσυστολές και τους μονότονους τελεστές.

Ορισμός 1.2.1: Έστω $(X, \|\cdot\|)$ χώρος με νόρμα. Μια συνάρτηση $H : X \to X$ καλείται *ψευδοσυστολή*, εαν υπάρχουν $x^* \in X$ και $\beta \in [0,1)$ τέτοια ώστε

$$\|Hx - x^*\| \leq \beta \|x - x^*\| \tag{1.2.4}$$

για κάθε $x \in X$.

Συνήθως, οι τελεστές H δε θα είναι συστολές ως προς την ευκλείδεια νόρμα του $\mathbb{R}^N$, αλλά ως προς κάποια ισοδύναμη νόρμα $\|\cdot\|_\xi$. Σημειώνουμε επίσης ότι κάθε ψευδοσυστολή έχει ως μοναδικό σταθερό της σημείο το x^*.

Πρόταση 1.2.2: *Έστω $(r_n)_n$ η ακολουθία που ορίζεται από την αναδρομική σχέση*

$$r_{n+1} = (1-\gamma_n)r_n + \gamma_n(H_n r_n + w_n + u_n), \tag{1.2.5}$$

όπου

(a) η ακολουθία $(\gamma_n)_n$ είναι τέτοια ώστε $\sum_{n=1}^{\infty} \gamma_n(i) = \infty$ και $\sum_{n=1}^{\infty} \gamma_n(i)^2 < \infty$ για κάθε $i = 1, \ldots, N$.

(b) Η ακολουθία $(w_n)_n$ έχει την ιδιότητα ότι

$$\mathbb{E}[w_n(i) \mid \mathcal{F}_n] = 0 \quad \text{και} \quad \mathbb{E}\left[w_n(i)^2 \mid \mathcal{F}_n\right] \leq A + B\|r_n\|^2.$$

(c) Κάθε H_n είναι ψευδοσυστολή ως προς την ίδια νόρμα $\|\cdot\|_\xi$, με το ίδιο σταθερό σημείο r^* και την ίδια σταθερά $\beta \in [0,1)$.

(d) Υπάρχει ακολουθία μη-αρνητικών τυχαίων μεταβλητών $(\theta_n)_n$ η οποία συγκλίνει στο μηδέν σχεδόν παντού, τέτοια ώστε

$$\|u_n\|_\infty \leq \theta_n\left(1 + \|r_n\|_\xi\right)$$

για κάθε $n \in \mathbb{N}$.

Τότε η $(r_n)_n$ συγκλίνει στο r^* σχεδόν παντού.

Πρόταση 1.2.3: *Έστω $(r_n)_n$ η ακολουθία που ορίζεται από την αναδρομική σχέση*

$$r_{n+1} = (1-\gamma_n)r_n + \gamma_n(Hr_n + w_n), \tag{1.2.6}$$

όπου

(a) η ακολουθία $(\gamma_n)_n$ είναι τέτοια ώστε $\sum_{n=1}^{\infty}\gamma_n(i) = \infty$ και $\sum_{n=1}^{\infty}\gamma_n(i)^2 < \infty$ για κάθε $i = 1,\dots,N$.

(b) Η ακολουθία $(w_n)_n$ έχει την ιδιότητα ότι

$$\mathbb{E}[w_n(i) \mid \mathcal{F}_n] = 0 \quad \text{και} \quad \mathbb{E}\left[w_n(i)^2 \mid \mathcal{F}_n\right] \le A + B\|r_n\|^2.$$

(c) Για τον τελεστή H ισχύει ότι

(i) είναι μονότονος, δηλαδή $Hx \le Hy$ για κάθε $x \le y$.

(ii) Για κάθε $\lambda > 0$ και $r \in \mathbb{R}^N$, ισχύει ότι: $Hr - \lambda e \le H(r-\lambda e) \le H(r+\lambda e) \le Hr + \lambda e$, όπου $e = (1,\dots,1)$.

(iii) Έχει μοναδικό σταθερό σημείο, $Hr^* = r^*$.

Εαν η $(r_n)_n$ είναι φραγμένη σχεδόν παντού, τότε συγκλίνει στο r^* σχεδόν παντού.

1.2.3 *Q-Μάθηση*

Ο δυναμικός προραμματισμός αποτελεί εναν από τους σημαντικότερους κλάδους της μαθηματικής βελτιστοποίησης και το αντικείμενο μελέτης του είναι προβλήματα στα οποία η λήψη αποφάσεων γίνεται ακολουθιακά.

Κεντρικό ρόλο στα προβλήματα δυναμικού προγραμματισμού, διαδραματίζει η εξίσωση του Bellman, η οποία αποτελεί μια συναρτησιακή εξίσωση για τη συνάρτηση βέλτιστης τιμής J^*:

$$J^*(i) = \min_{a\in A(i)}\left\{\sum_{j\in S} p_{ij}(a)\left(c(i,a,j) + J^*(j)\right)\right\}. \tag{1.2.7}$$

Η επίλυση της εξίσωσης αυτής είναι άμεσα συνυφασμένη και με τη λύση του υπό μελέτη προβλήματος. Για απλά προβλήματα, η εξίσωση (1.2.7) γίνεται να λυθεί αλγεβρικά, αλλά εν γένει αυτό δεν είναι εφικτό. Για τον λόγο αυτό, ακόμα και στη κλασσική θεωρία του δυναμικού προγραμματισμού, είναι σημαντική η εύρεση μεθόδων για την προσεγιστική επίλυσή της. Ένα παράδειγμα τέτοιας μεθόδου είναι η *μέθοδος των διαδοχικών προσεγγίσεων*, σύμφωνα με την οποία ξεκινάμε με κάποια αυθαίρετη συνάρτηση J_0 και σε κάθε στάδιο ορίζουμε

$$J_{n+1}(i) = \min_{a\in A(i)}\left\{\sum_{j\in S} p_{ij}(a)\left(c(i,a,j) + J_n(j)\right)\right\}. \tag{1.2.8}$$

Η σύγκλιση της ακολουθίας $(J_n)_n$ στην J^* εξασφαλίζεται μέσω του θεωρήματος σταθερού σημείου του Banach, καθώς ο τελεστής $T : C(S) \to C(S)$, που ορίζεται ώς

$$(Tf)(i) := \min_{a \in A(i)} \left\{ \sum_{j \in S} p_{ij}(a) \left(c(i,a,j) + f(j) \right) \right\}, \ f \in C(S), \ i \in S,$$

αποτελεί συστολή.

Η Q-Μάθηση μπορεί να θεωρηθεί σαν μια επέκταση αυτού του αποτελέσματος, για την επίλυση προβλημάτων όπου είναι αδύνατη η χρησιμοποίησή του, λόγω έλλειψης απαιτούμενων πληροφοριών. Στην κλασσική περίπτωση, οι πιθανότητες μετάβασης $p_{ij}(a)$ καθώς και τα κόστη $c(i,a,j)$ είναι εκ των προτέρων γνωστά, έτσι ο υπολογισμός της έκφρασης στη σχέση (1.2.8) είναι εφικτός. Στο πλαίσιο της Q-Μάθησης, οι τιμές αυτές είναι άγνωστες. Αντίθετα, υποθέτουμε ότι μπορούμε να προσομοιώνουμε τις μεταβάσεις της αλυσίδας οι οποίες περιγράφονται από τις κατανομές $p_{ij}(a)$ και να παρατηρούμε τα κόστη που επιφέρει κάθε μετάβαση. Έτσι, ενώ δε γνωρίζουμε τις ακριβείς παραμέτρους του προβλήματος, διαθέτουμε έναν έμμεσο τρόπο να εξάγουμε συμπεράσματα για αυτές.

Ο αλγόριθμος της Q-Μάθησης διατυπώθηκε από τον Chris Watkins στη διδακτορική διατριβή του [Wat89] και αποτελέι ένα συνδυασμό της μεθόδου διαδοχικών προσεγγίσεων με τη μέθοδο της στοχαστικής προσέγγισης των Robbins-Monro. Το υπό μελέτη πρόβλημα ανάγεται στην επίλυση της συναρτησιακής εξίσωσης

$$Q(i,a) := (1-\gamma)Q(i,a) + \gamma \sum_{j=0}^{N} p_{ij}(a) \left(c(i,a,j) + \min_{b \in A(j)} Q(j,b) \right) \qquad (1.2.9)$$

για κάθε κατάσταση i και απόφαση $a \in A(i)$. Καθώς δε γνωρίζουμε τις πιθανότητες μετάβασης $p_{ij}(a)$, η αναμενόμενη τιμή που εμφανίζεται στη σχέση (1.2.9), αντικαθίσταται από μία τιμή j η οποία προσομοιώνεται από την κατανομή $p_{i,\cdot}(a)$. Ομοίως, το $c(i,a,j)$ δεν είναι εκ των προτέρων γνωστό, αλλά παρατηρείται κατά την πραγματοποίηση της εν λόγω μετάβασης. Τέλος, η σταθερά γ αντικαθίσταται από ακολουθίες που συγκλίνουν στο μηδέν αρκετά αργά, όπως και στον αλγόριθμο των Robbins-Monro:

Θεώρημα 1.2.4: [BT96] *Θεωρούμε την ακολουθία $(Q_n)_n$ που ορίζεται αναδρομικα από τη σχέση*

$$Q_{n+1}(i,a) = \left(1 - \gamma_n(i,a)\right) Q_n(i,a) + \gamma_n(i,a) \left(c(i,a,j) + \min_{b \in A(j)} Q_n(j,b) \right), \quad (1.2.10)$$

όπου σε κάθε βήμα, η τιμή j έχει προσομοιωθεί από την κατανομή $p_{i,\cdot}(a)$ και η ακολουθία $(\gamma_n)_n$ είναι τέτοια ώστε $\sum_{n=0}^{\infty} \gamma_n(i,a) = \infty$ και $\sum_{n=0}^{\infty} \gamma_n(i,a)^2 < \infty$ για κάθε $i = 1, \ldots, N$ και $a \in A(i)$. Κάτω από κατάλληλες υποθέσεις, $Q_n(i,a) \to Q^*(i,a)$ για κάθε i, $a \in A(i)$ σχεδόν παντού, όπου Q^* είναι ο βέλτιστος Q- παράγοντας, δηλαδή η λύση της εξίσωσης (1.2.9).

Η ιδέα της απόδειξης είναι να οριστεί κατάλληλα ένας τελεστής H, όπως και στην απόδειξη της σύγκλισης της μεθόδου διαδοχικών προσεγγίσεων, το σταθερό σημείο του οποίου θα είναι ο ζητούμενος βέλτιστος Q-παράγοντας και ο οποίος θα εμπίπτει σε κάποια από τα θεωρήματα στοχαστικής προσέγγισης της προηγούμενης παραγράφου. Συγκεκριμένα, ο τελεστής $H : C(\tilde{S}) \to C(\tilde{S})$ που ορίζεται ως

$$(HQ)(i,a) = \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b\in A(j)} Q(j,b)\right). \tag{1.2.11}$$

για Q στο $C(\tilde{S})$, προκύπτει είτε ότι είναι συστολή ως προς κάποια κατάλληλη νόρμα, είτε ότι είναι μονότονος, έτσι τα αποτελέσματα της προηγούμενης παραγράφου μπορούν να εφαρμοστούν.

1.3 *Multi-armed bandits*

Το πλαίσιο των multi-armed bandits αποτελεί ένα από τα πιο αντιπροσωπευτικά και εύληπτα παραδείγματα του κλάδου της ενισχυτικής μάθησης (reinforcement learning). Μελετήθηκαν για πρώτη φορά από τον Herbert Robbins στις αρχές τις δεκαετίας του '50 και εξακολουθούν να αποτελοούν κεντρικό πεδίο έρευνας, τόσο από θεωρητική άποψη, όσο και στο κομμάτι των εφαρμογών τους.

Στην απλούστερη μορφή του, το πρόβλημα που θα μας απασχολήσει είναι το εξής: Έχουμε τη δυνατότητα να τραβάμε παρατηρήσεις από δύο διαφορετικούς πληθυσμούς A και B, οι οποίοι χαρακτηρίζονται από τις κατανομές F_A και F_B, με αναμενόμενες τιμές a και b αντίστοιχα. Σε κάθε γύρο, επιλέγουμε έναν εκ των δύο πληθυσμών, τραβάμε μια παρατήρηση x σύμφωνα με την αντίστοιχη κατανομή, την οποία και εισπράττουμε σαν κέρδος. Ο στόχος μας είναι να βρούμε μια στρατηγική επιλογής του δείγματος $x_1, \ldots, x_n$ ούτως ώστε να μεγιστοποιήσουμε το αναμενόμενο κέρδος $S_n = x_1 + \ldots + x_n$.

Εαν γνωρίζαμε εξ' αρχής τις τιμές των a και b, τότε το πρόβλημα θα ήταν τετριμμένο, αφού σε κάθε γύρο θα επιλέγαμε τον πληθυσμό με την μεγαλύτερη. Από τη στιγμή όμως που οι τιμές αυτές είναι άγνωστες, τότε σε κάθε γύρο το μέσο κέρδος μας θα έχει τη μορφή

$$\mathbb{E}[S_n/n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{k_n}{n}a + \frac{n-k_n}{n}b,$$

όπου k_n ο αριθμός δειγμάτων από τον πληθυσμό A μετά από n γύρους. Δηλαδή, το μέσο κέρδος $\mathbb{E}[S_n/n]$ θα αποτελεί, εν γένει, έναν κυρτό συνδυασμό των a και b.

Η ιδιαιτερότητα του προβλήματος συνίσταται στο ότι για να το προσεγγίσουμε σωστά, είναι απαραίτητο να χρησιμοποιήσουμε ταυτόχρονα δύο αντικρουόμενες στρατηγικές. Από τη μία μεριά, πρέπει να τραβήξουμε αρκετές τιμές και από τους δύο πληθυσμούς ώστε να έχουμε μια καλή εκτίμηση του ποιος εκ των δύο είναι ο καλύτερος. Αυτό σημαίνει ότι είμαστε διατεθειμένοι να επιλέξουμε τον λάθος

πληθυσμό αρκετές φορές ούτως ώστε να μάθουμε την αναμενόμενη τιμή του. Από την άλλη, όταν είμαστε αρκετά βέβαιοι για το ποιος είναι ο καλύτερος, πρέπει να τραβάμε ολοένα περισσότερες παρατηρήσεις από αυτόν τον πληθυσμό, ώστε να μεγιστοποιήσουμε το αναμενόμενο κέρδος μας. Η πρώτη στρατηγική ονομάζεται *εξερεύνηση*, ενώ η δεύτερη *εκμετάλλευση*, και η εύρεση αλγορίθμων που επιτυγχάνουν τη σωστή εξισορρόπηση ανάμεσά τους είναι το αντικείμενο του Κεφαλαίου 4.

Υπάρχει ένας πολύ απλός αλγόριθμος, ο οποίος διατυπωθηκε από τον ίδιο τον Robbins [Rob52] το 1952, και επιτυγχάνει ασυμπτωτικά το μέγιστο δυνατό αναμενόμενο κέρδος. Σύμφωνα με αυτόν, επιλέγονται δύο ξένα, άπειρα υποσύνολα φυσικών αριθμών J_A και J_B με μηδενική πυκνότητα.[2] Όταν ο γύρος n στον οποίο βρισκόμαστε, ανήκει στο J_A, επιλέγουμε πάντα τον πληθυσμό A. Αντίστοιχα επιλέγουμε τον B πληθυσμό όταν $n \in J_B$. Σε όλους τους υπόλοιπους γύρους, επιλέγουμε τον πληθυσμό ο οποίος έχει τον μεγαλύτερο δειγματικό μέσο, σύμφωνα με τις παρατηρήσεις που έχουν ληφθεί εως εκείνη τη στιγμή.

Καθώς τα σύνολα J_A και J_B έχουν πυκνότητα μηδέν, οι παρατηρήσεις που λαμβάνουμε επάνω τους δεν έχουν κάποια επίδραση στο μέσο κέρδος μας ασυμπτωτικά. Επομένως, πάνω σε αυτά τα σύνολα μπορούμε να εξερευνούμε όσο θέλουμε, χωρίς αρνητικές συνέπειες. Επιπλέον, εφόσον και τα δύο αυτά σύνολα είναι άπειρα, η στρατηγική μας είναι σίγουρο ότι θα επιλέγει πάντα άπειρες το πλήθος παρατηρήσεις και από τους δύο πληθυσμούς. Από τον Ισχυρό Νόμο των Μεγάλων Αριθμών, οι δειγματικοί μέσοι των δύο πληθυσμών θα συγκλίνουν στους πραγματικούς μέσους με πιθανότητα ένα.

Άρα, από κάποιον γύρο και έπειτα, οι δειγματικοί μέσοι των δύο πληθυσμών θα είναι τόσο κοντά στους πραγματικούς μέσους, ώστε επιλέγοντας τον μεγαλύτερο δειγματικό μέσο, ουσιαστικά επιλέγουμε και τον καλύτερο πληθυσμό. Έτσι, η στρατηγική μας από ένα γύρο και μετά, θα επιλέγει μονίμως τον καλύτερο πληθυσμό, εκτός από τους γύρους που ανήκουν στο σύνολο $J_A \cup J_B$, οι οποίοι όμως είναι τόσο αραιοί που δεν επηρεάζουν το αναμενόμενο κέρδος μας ασυμπτωτικά.

1.3.1 *Ασυμπτωτικά βέλτιστα κάτω φράγματα*

Σε πραγματικά προβλήματα, όπου δεν συνεχίζουμε να παίζουμε επ' άπειρον, είναι σημαντικό όχι απλά να εξασφαλίσουμε την επίτευξη του στόχου ασυμπτωτικά, αλλά και όσο το δυνατόν γρηγορότερα. Όμως, ενώ η εύρεση αλγορίθμων που επιτυγχάνουν το μέγιστο δυνατό αναμενόμενο κέρδος ασυμπτωτικά ήταν μια απλή εφαρμογή του INMA, το ερώτημα του πόσο γρήγορα μπορεί να επιτευχθεί αυτό το κέρδος αποδείχθηκε πολύ δυσκολότερο και απαντήθηκε 30 χρόνια αργότερα από τους Tze Leung Lai και Herbert Robbins [LR85].

[2] Για παράδειγμα μπορούμε να επιλέξουμε $J_A = \{n^2 : n \in \mathbb{N}\}$ και $J_B = \{n^2 + 1 : n \in \mathbb{N}\}$.

Θα θεωρήσουμε την πιο γενική περίπτωση, όπου έχουμε k το πλήθος πληθυσμούς με αναμενόμενες τιμές $a_1, \dots, a_k$. Για κάθε στρατηγική $\phi : \mathbb{N} \to \{1, \dots, k\}$ ορίζουμε την *απώλεια* (regret) μετά από n γύρους, ως

$$R_n(\phi) = n \max\{a_1, \dots, a_k\} - \mathbb{E}[S_n]. \tag{1.3.1}$$

Οι Lai και Robbins έδειξαν ότι στην περίπτωση που οι κατανομές των κερδών ικανοποιούν κάποιες συγκεκριμένες υποθέσεις, η απώλεια οποιουδήποτε αλγορίθμου δε μπορεί να αυξάνει πιο αργά από $o(\ln n)$. Επιπλέον, κατασκεύασαν έναν αλγόριθμο ο οποίος επιτυγχάνει αυτό το κάτω φράγμα, και άρα συμπεριφέρεται βέλτιστα όσον αφορά στην ταχύτητα σύγκλισης.

Όλες οι κατανομές στο συγκεκριμένο θεώρημα είναι μονοπαραμετρικές $f(x;\theta)$ με $\theta \in \Theta \subseteq \mathbb{R}$, και το Θ με τη σειρά του ικανοποιεί κάποια αξιώματα που εξασφαλίζουν ότι θα έχει μια σχετικά πλούσια δομή. Επίσης, διατυπώνονται κάποιες υποθέσεις συνέχειας της μετρικής Kullback-Leibler. Όλες αυτές οι υποθέσεις, μπορεί εκ πρώτης όψεως να φαίνονται αντιδιαισθητικές, αλλά στην πραγματικότητα εξασφαλίζουν ότι το πρόβλημα είναι διατυπωμένο σε ρεαλιστικά πλαίσια και επίσης αποφεύγουν τετριμμένες απαντήσεις.

Για την εύρεση του κάτω φράγματος $o(\ln n)$, οι Lai και Robbins χρησιμοποιούν το εξής βασικό επιχείρημα: Αν υποθέσουμε ότι έχουμε στα χέρια μας έναν αλγόριθμο ο οποίος συγκλίνει γρήγορα για όλες τις δυνατές κατανομές, δηλ. για όλα τα δυνατά $\theta \in \Theta$, τότε αυτή ακριβώς η ισχυρή του ιδιότητα, στην πραγματικότητα λειτουργεί σε βάρος του: Επιλέγοντας με προσοχή συγκεκριμένες τιμές τις παραμέτρου θ, στις οποίες ο αλγόριθμος αναγκαστικά θα πρέπει να συγκλίνει γρήγορα, καταφέρνουν και παίρνουν αντίστροφα φράγματα. Το ηθικό δίδαγμα της απόδειξης, είναι ότι εαν έχουμε στην κατοχή μας έναν αλγόριθμο ο οποίος συγκλίνει γρήγορα για όλα τα $\theta \in \Theta$, τότε αναγκαστικά θα υπάρχουν κάποια θ στα οποία, θα συγκλίνει μεν γρήγορα, αλλά δε θα συγκλίνει ταχύτατα.

Για την κατασκευή του αλγορίθμου που το υλοποιεί, χρησιμοποιούν την αρχή της *αισιοδοξίας εν όψει αβεβαιότητας* (optimism in the face of uncertainty). Η κύρια διαφορά σε σχέση με τον πολύ απλό αλγόριθμο που αναφέραμε στην προηγούμενο παράγραφο, είναι ότι σε κάθε γύρο δε συγκρίνονται οι δειγματικοί μέσοι των διαφόρων πληθυσμών. Αντίθετα, για τους πληθυσμούς που δεν έχουν επιλεγεί αναλογικά αρκετές φορές, κατασκευάζεται ένα διάστημα εμπιστοσύνης για τον μέσο, και οι ίδιοι οι πληθυσμοί αντιπροσωπεύονται όχι από τον δειγματικό τους μέσο, αλλά από το άνω άκρο του διαστήματος εμπιστοσύνης. Με αυτή την πρακτική, ο αλγόριθμος δίνει ένα παραπάνω κίνητρο προς εξερεύνηση, προσφέροντας επιπλέον ευκαιρίες σε πληθυσμούς που δείχνουν υποσχόμενοι, αλλά ίσως να έχουν μικρούς δειγματικούς μέσους λόγω τυχαιότητας.

1.3.2 *Ο αλγόριθμος UCB*

Η κατασκευή των διαστημάτων εμπιστοσύνης στη δουλειά των Lai και Robbins είναι ιδιαίτερα λεπτή υπόθεση. Στο άρθρο τους, αναφέρονται τα αξιώματα που τα

διαστήματα εμπιστοσύνης θα πρέπει να ικανοποιούν, και δίνονται παραδείγματα για συγκεκριμένες κατανομές, αλλά εν γένει, δεν παρέχεται κάποιος κανόνας κατασκευής τους. Ακόμα όμως και όταν τα διαστήματα είναι γνωστά, ο υπολογισμός τους είναι αρκετά απαιτητικός.

Το πρόβλημα αυτό προσπαθεί να αντιμετωπίσει ο αλγόριθμος Upper Confidence Bound (UCB) των Auer, Cesa-Bianchi και Fischer [ACBF02], στον οποίον τα διαστήματα εμπιστοσύνης έχουν την απλούστατη έκφραση

$$\text{Διάστημα Εμπιστοσύνης } j\text{-πληθυσμού} = \overline{x}_{j,n_j} \pm \sqrt{\frac{3\ln n}{2n_j}}, \tag{1.3.2}$$

όπου $\overline{x}_{j,n_j}$ είναι ο δειγματικός μέσως του j-πληθυσμού και n_j το πλήθος φορών που επιλέχθηκε κατά τους πρώτους n γύρους. Επιπλέον, η μόνη υπόθεση για τις κατανομές των κερδών είναι να λαμβάνουν τιμές στο $[0,1]$.

Το τίμημα που πληρώνουμε είναι ότι, ενώ μεν επιτυγχάνεται και πάλι ταχύτητα σύγκλισης $o(\ln n)$, η σταθερά είναι χειρότερη από τη βέλτιστη δυνατή. Τροποποιώντας όμως κατάλληλα τον αλγόριθμο αυτόν, η σταθερά μπορεί να πλησιάσει αυθαίρετα κοντά στη βέλτιστη σταθερά. Αυτό υλοποιείται στον αλγόριθμο Upper Confidence Bounds with epochs, όπου σε κάθε γύρο, ο πληθυσμός που επιλέγεται, δεν παίζεται μόνο μία φορά, αλλά ολοένα και περισσότερες φορές.

Part I

THEORETICAL FOUNDATIONS

2

Approximation Capabilities of Neural Networks

Neural Networks have exhibited tremendous results during the past decades, offering approximate solutions to problems often conceived intractable. Despite the fact that they were firstly conceived in the mid '40s, the systematic study of their approximation capabilities emerged relatively recently. It relied on tools from abstract mathematics, like functional analysis and probability theory, and progressively reached deeper into their foundations.

In Section 2.1 we describe the mathematical formulation of the problem, and in Section 2.2 we present Cybenko's approximation theorem, which is the first major relevant theorem, proving that continuous sigmoidal activation functions are universal approximators. In Section 2.3 we investigate if the sigmoidal property of the activation function is really necessary for the approximation property to hold. Leshno, Lin, Pinkus and Schocken answered this question in the negative, providing an impressive and simple characterization of the activation functions that are universal approximators. In Section 2.4 we present a constructive approximation method in L_2, based on the work of Kwok and Yeung. Lastly, Section 2.5 is devoted to an alternative method of obtaining approximation results, based on the Kolmogorov - Arnold Representation Theorem.

2.1 *The universal approximation property*

In Approximation Theory, one is usually trying to approximate functions using simpler ones. Probably the most famous relevant result is the Stone - Weierstrass Theorem, which asserts that one can approximate continuous functions using polynomials. Artificial Neural Networks are architectures that are used in this setting, however, the "simpler" functions that a neural network uses in order to approximate more complex ones, are not polynomials, but have another very specific form [BT96, p. 64]: Any function s created by a neural network has the form $s(x) = Tx$, where

$$T = A_{m+1} S_m A_m \cdots A_2 S_1 A_1$$

is an operator consisting of sequential compositions of two types of transformations; the affine transformations A_i and the sigmoidal ones S_i.

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is the function we are trying to approximate. Let $m \in \mathbb{N}$ and consider a sequence of natural numbers $d_0 = n$, $d_1, \ldots, d_{m+1}$. For each

$i = 1,\dots,m+1$, let $A_i : \mathbb{R}^{d_{i-1}} \to \mathbb{R}^{d_i}$ be an affine transformation.[1] Additionally, suppose that $\sigma_1,\dots,\sigma_m : \mathbb{R} \to \mathbb{R}$ are real functions (not necessarily linear), and for every $i = 1,\dots,m$, every $d \in \mathbb{N}$ and $z \in \mathbb{R}^d$, we define $S_i z$ to be the vector $S_i z = [\sigma_i(z_1),\dots,\sigma_i(z_d)]^\intercal$. Specifically, S_i applies the function σ_i to every element of the vector z. Note here that S_i is a (not necessarily linear) function that respects the input dimension, that is, $S_i : \mathbb{R}^d \to \mathbb{R}^d$ for every d.

Under this notation, given an input $x \in \mathbb{R}^n$, the output $s(x)$ of a neural network with m *layers* and having $\sigma_1,\dots,\sigma_m$ as its *activation functions*, is:

$$s(x) = A_{m+1} S_m A_m \cdots A_2 S_1 A_1 x. \tag{2.1.1}$$

The activation functions σ_i, are most commonly sigmoidal functions, that is, functions having the property that $\lim_{t\to-\infty}\sigma(t) = 0$ and $\lim_{t\to+\infty}\sigma(t) = 1$. A single layered network ($m = 1$) has an output of the form

$$s(x) = A_2 S_1 A_1 x.$$

In this particular case, the set of all possible functions s which can be generated by a neural network, having σ as an activation function, is:

$$\Sigma_n(\sigma) = \Big\{ s : \mathbb{R}^n \to \mathbb{R} : s(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j) : N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \Big\}.$$

The elements $w_j^\intercal x + \theta_j$ correspond to the first affine transformation, $A_1 : \mathbb{R}^n \to \mathbb{R}^N$, applied on x, namely

$$A_1 x = (w_1^\intercal x + \theta_1, \dots, w_N^\intercal x + \theta_N).$$

For the next step, on each of the coordinates of this vector, the activation function σ is applied, producing the vector

$$S_1 A_1 x = (\sigma(w_1^\intercal x + \theta_1), \dots, \sigma(w_N^\intercal x + \theta_N))$$

Finally, the (linear) operator $A_2 : \mathbb{R}^N \to \mathbb{R}$, defined as $A_2(x) = \sum_{i=1}^N a_i x_i$ for $x \in \mathbb{R}^N$, is applied, to produce the final linear combination $\sum_{j=1}^N a_j \sigma(w_j^\intercal x + \theta_j)$. [2]

[1] A function $A : X \to Y$ between two vector spaces is called *affine*, when it is the translation of a linear operator. This means that there exist $T : X \to Y$ linear and $y_0 \in Y$, such that $A(x) = T(x) + y_0$ for every $x \in X$. Equivalently, A is affine if $A\left(\sum_{i=1}^n \lambda_i x_i\right) = \sum_{i=1}^n \lambda_i A(x_i)$ for every $n \in \mathbb{N}$, $x_i \in X$ and $\lambda_i \in \mathbb{R}$ such that $\sum_{i=1}^n \lambda_i = 1$. Linear operators are special cases of affine transformations; this can readily be seen by picking $y_0 = 0$ in the first definition given above, or by observing that for linear operators T, the identity $T\left(\sum_{i=1}^n \lambda_i x_i\right) = \sum_{i=1}^n \lambda_i T(x_i)$ holds for any $n \in \mathbb{N}$ and any choice of λ_i.

[2] One could observe here that from the final sum, the constant part of the affine transformation A_2 appears to be missing. However, by picking $w_j = 0$ and θ_j such that $\sigma(\theta_j) \neq 0$, we can make sure that the constant function will appear in this last combination, so we can use a linear combination, instead of an affine one, for simplicity.

The use of neural networks in approximation problems has been proven to be extremely successful. Of course, one cannot expect that any function f will be exactly equal to some function of the form (2.1.1), however, one could hope that it would be possible to approximate any given f arbitrarily close, using functions of this form. This is indeed the case, and is achievable for a variety of activation functions. Activation functions which have this property, are called *universal approximators*. In what follows, we will focus on single layered networks,[3] since if we manage to show that an activation function σ is a universal approximator for such a network, then clearly the same would hold for multilayered networks as well.

We should point out here that although approximation theorems assert that we can always approximate any unknown function f, they are not designed to provide a rule of how to do so. In real life applications, the researcher is usually given a vector of observations $(x_i, f(x_i))_i$, or more realistically $(x_i, y_i)_i$ where $y_i = f(x_i) + \varepsilon_i$ for some noise ε_i, and just based on it, he tries to determine the appropriate values of the parameters w_j, θ_j, as well as the activation function σ, so as the corresponding predicted values of his model $(s(x_i))_i$ are "close" to the observed ones $(y_i)_i$, expecting that the resulting s will be acceptably close to f on the whole domain. The importance of the approximation theorems in this context, is that they ascertain that his efforts are not in vain: parameters for which s is sufficiently close to the real f do exist, as long as he chooses a universal approximator as an activation function.

2.2 *Cybenko's approximation theorem*

One of the first and most important relevant theorems, belongs to George Cybenko [Cyb89] who, in 1989, proved that continuous sigmoidal activation functions are universal approximators. The proof is a beautiful application of the Hahn-Banach and Riesz Representation theorems. The reader who is unfamiliar with the terminology, or the results, may need to consult Appendix A.2 for a basic treatment, or his favorite functional analysis book for a more thorough one.

We use $I_n = [0,1]^n$ to denote the n-dimensional cube and by $C(I_n)$ we denote the Banach space of continuous functions on I_n, equipped with the supremum norm $\|f\| = \|f\|_\infty = \sup\{|f(x)| : x \in I_n\}$. Since I_n is compact, the Riesz Representation Theorem [AB06, Corollary 14.15] applies and asserts that the dual space of $C(I_n)$ can be identified as the space of finite, signed Borel measures on I_n, denoted as $M(I_n)$.

[3] Intuitively, one expects that a multilayered network will be at least as good of an aproximator as a single layered one, as long as the additional A_i's and S_i's are chosen properly. For example, if the output of the network is $s(x) = A_{m+1} \dots A_3 S_2 A_2 S_1 A_1 x$, then by using $\sigma(x) = x$ and $A = I$ to define all the additional operators, we have that $A_{m+1} S_m A_m \dots A_3 S_2$ is the identity operator, thus our network is able to reproduce the single layered network defined by $A_2 S_2 A_1$.

Definition 2.2.1: A function $\sigma : \mathbb{R} \to \mathbb{R}$ is called *sigmoidal* if

$$\sigma(t) \longrightarrow \begin{cases} 0, & \text{as } t \to -\infty, \\ 1, & \text{as } t \to +\infty. \end{cases} \tag{2.2.1}$$

Given a sigmoidal function σ, the set

$$\begin{aligned} \Sigma_n(\sigma) = \Big\{ f : I_n \to \mathbb{R} : f(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j) : \\ N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \Big\} \\ = \operatorname{span} \big\{ f : f(x) = \sigma\left(w^\intercal x + \theta\right) \text{ for } w \in \mathbb{R}^n, \theta \in \mathbb{R} \big\}. \end{aligned}$$

contains all the possible functions that can be generated by a single layered neural network, having σ as an activation function. Cybenko's theorem states that $\Sigma_n(\sigma)$ is dense[4] in $C(I_n)$ for continuous sigmoidal functions σ.

Definition 2.2.2: Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a real function and $\mu \in M(I_n)$ be a measure. We say that σ is *discriminatory* for μ if

$$\int_{I_n} \sigma(w^\intercal x + \theta) d\mu(x) = 0 \tag{2.2.2}$$

for every $w \in \mathbb{R}^n$ and every $\theta \in \mathbb{R}$ implies that $\mu \equiv 0$.

The proof idea[5] is to show that $\Sigma_n(\sigma)$ is dense in $C(I_n)$ whenever σ is discriminatory, and then prove that sigmoidal functions share the discriminatory property. It is based on two fundamental theorems in functional analysis, the Hahn-Banach and the Riesz Reperesentation Theorems.

Theorem 2.2.3 (Hahn-Banach): [Arg04, Proposition 5.7] *Let X be a normed space and Y be a closed proper subspace of X. Then, there exists a nonzero $x^* \in X^*$ such that $x^*(y) = 0$ for every $y \in Y$.*

Theorem 2.2.4 (Riesz Representation): [AB06, Theorem 14.16] *Let X be a compact metrizable space and $F \in C(X)^*$ be a bounded, linear functional on $C(X)$. Then, there exists a unique, finite, signed Borel measure μ on X, such that $F(f) = \int_X f d\mu$ for every $f \in C(X)$.*

[4] A subset $D \subseteq X$ of a metric space (X, ρ) is dense if $\overline{D} = X$. There are several equivalent reformulations of this definition, the most useful of which is that for every $\varepsilon > 0$ and every $x \in X$ there exists some $d \in D$ with $\rho(d, x) < \varepsilon$. In our setting, $\Sigma_n(\sigma)$ being dense in $C(I_n)$ means that for every $\varepsilon > 0$ and $f \in C(I_n)$, there exists some $g \in \Sigma_n$ with $\|f - g\|_\infty < \varepsilon$. Simply put, g approximates f well enough.

[5] One could be tempted to use the Stone-Weierstrass Theorem to conclude that $\Sigma_n(\sigma)$ is dense. Unfortunately, the Stone-Weierstrass Theorem is not applicable here without some modifications. We postpone this discussion until p. 28.

Theorem 2.2.5: *Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a continuous and discriminatory function. Then, the set*

$$\Sigma_n(\sigma) = \left\{ f : I_n \to \mathbb{R} : f(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j) : \right.$$
$$\left. N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \right\} \tag{2.2.3}$$

is dense in $C(I_n)$.

Proof. The set $\Sigma_n(\sigma)$ is clearly a subspace of $C(I_n)$. Indeed, if $\lambda \in \mathbb{R}$ and $f(x) = \sum_{j=1}^{N} a_j \sigma(y_j^\intercal x + \theta_j) \in \Sigma_n(\sigma)$, then

$$\lambda f(x) = \sum_{j=1}^{N} \lambda a_j \sigma(y_j^\intercal x + \theta_j) = \sum_{j=1}^{N} a_j' \sigma(y_j^\intercal x + \theta_j)$$

for $a_j' = \lambda a_j \in \mathbb{R}$, thus belongs to $\Sigma_n(\sigma)$. Similarly, if

$$f(x) = \sum_{j=1}^{N} a_j \sigma(y_j^\intercal x + \theta_j) \text{ and } g(x) = \sum_{j=1}^{M} b_j \sigma(z_j^\intercal x + k_j)$$

belong to $\Sigma_n(\sigma)$, their sum can be written as

$$f(x) + g(x) = \sum_{j=1}^{N+M} a_j \sigma(y_j^\intercal x + \theta_j) \in \Sigma(\sigma),$$

where we set $a_{N+i} = b_i$, $y_{N+i}^\intercal = z_i^\intercal$ and $\theta_{N+i} = k_i$ for every $i = 1, \ldots, M$. Alternatively, one could simply observe that $\Sigma_n(\sigma)$ is just the linear subspace of $C(I_n)$ generated by the functions of the form $f(x) = \sigma(w^\intercal x + \theta)$.

Suppose that $\Sigma_n(\sigma)$ is not dense in $C(I_n)$. Then, $Y = \overline{\Sigma_n(\sigma)}$ is a proper closed subspace of $C(I_n)$. By the Hahn-Banach Theorem, there exists some nonzero, bounded linear functional x^* on $C(I_n)$, such that $x^*(f) = 0$ for every $f \in Y$. By the Riesz Representation Theorem, there exists some finite, signed, nonzero Borel measure $\mu \in M(I_n)$, such that

$$x^*(f) = \int_{I_n} f(x) d\mu(x)$$

for every $f \in C(I_n)$. However, since x^* is zero on Y, this implies that for every $f \in \Sigma_n(\sigma) \subseteq Y$, $\int_{I_n} f(x) d\mu(x) = 0$, which combined with the discriminatory property of σ, yields that $\mu \equiv 0$, a contradiction. Therefore $\Sigma_n(\sigma)$ must be dense in $C(I_n)$. ∎

Theorem 2.2.6: *Any bounded, measurable sigmoidal function σ is discriminatory.*

Proof. Let σ be a bounded, measurable sigmoidal function and suppose that for every $w \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$, $\int_{I_n} \sigma(w^\intercal x + \theta) d\mu(x) = 0$ holds. We fix $w \in \mathbb{R}^n$ and $\theta, \phi \in$

$\mathbb{R}$ and define a sequence of real functions $(\sigma_k)_k$ on I_n as follows: For every $k \in \mathbb{N}$, let $\sigma_k(x) = \sigma(k(w^\intercal x + \theta) + \phi)$. We have that

$$\lim_{k\to\infty} \sigma_k(x) = \gamma(x) := \begin{cases} 0, & \text{when } w^\intercal x + \theta < 0, \\ 1, & \text{when } w^\intercal x + \theta > 0, \\ \sigma(\phi), & \text{when } w^\intercal x + \theta = 0. \end{cases} \tag{2.2.4}$$

Since σ is bounded, there exists some $M > 0$ such that $|\sigma_k(x)| \leq M$ for every $k \in \mathbb{N}$ and $x \in I_n$. By applying Lebesgue's Dominated Convergence Theorem, which still holds for signed measures, we obtain that

$$0 = \lim_{k\to\infty} \int_{I_n} \sigma_k(x) d\mu(x) = \mu(H_{w,\theta}) + \sigma(\phi)\mu(K_{w,\theta}), \tag{2.2.5}$$

where $H_{w,\theta} = \{x : w^\intercal x + \theta > 0\}$ and $K_{w,\theta} = \{x : w^\intercal x + \theta = 0\}$ are the open half-spaces and hyperplanes, defined by the parameters w, θ respectively. Since relation (2.2.5) holds for every $\phi \in \mathbb{R}$, by letting $\phi \to -\infty$ we obtain that $\mu(H_{w,\theta}) = 0$ for every $w \in \mathbb{R}^n$ and every $\theta \in \mathbb{R}$. Similarly, by letting $\phi \to +\infty$ we conclude the same for every hyperplane $K_{w,\theta}$. [6] For some fixed $w \in \mathbb{R}^n$ we define the transformation $T_w : \mathbb{R}^n \to \mathbb{R}$ by $T_w(x) = w^\intercal x$ for $x \in \mathbb{R}^n$. Let also $\nu = \mu T^{-1}$ denote the pushforward measure on $\mathbb{R}$, defined as $\nu(A) = \mu(T^{-1}(A)) = \mu(\{x : w^\intercal x \in A\})$ for every Borel $A \subseteq \mathbb{R}$.

Consider the linear functional $F : L^1(\mathbb{R}, \nu) \to \mathbb{R}$ defined as

$$F(h) = \int_{I_n} h\left(w^\intercal x\right) d\mu(x) \quad \text{for } h \in L^1(\mathbb{R}, \nu).$$

According to the Change of Variables formula, [AB06, Theorem 13.46][7]

$$F(h) = \int_{I_n} h\left(w^\intercal x\right) d\mu(x) = \int_{\mathbb{R}} h(t) d\mu T^{-1}(t) = \int_{\mathbb{R}} h(t) d\nu(t) \tag{2.2.6}$$

and

$$|F(h)| = \left| \int_{\mathbb{R}} h(t) d\nu(t) \right| \leq \int_{\mathbb{R}} |h(t)| d|\nu|(t) =: \|h\|_{1,\nu}, \tag{2.2.7}$$

so F is also bounded. For every $\theta \in \mathbb{R}$, the function $h = I_{(\theta,\infty)}$ belongs to $L^1(\mathbb{R}, \nu)$, so

$$F(I_{(\theta,\infty)}) = \mu\left(\{x : w^\intercal x > \theta\right) = \mu(H_{w,\theta}) = 0. \tag{2.2.8}$$

[6] If μ was a positive measure, this would be enough to conclude that $\mu \equiv 0$, since I_n can be written as the union of two such half-spaces, $I_n = H_{w_0,\theta_0} \cup H_{-w_0,-\theta_0}$ for some arbitrary w_0 and θ_0. However, the measure μ is signed, so we need a more involved argument to deduce that μ is zero.

[7] Suppose that $T : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B})$ is a measurable function and $\nu = \mu T^{-1}$ is the pushforward measure on Y Then for any integrable function $f : Y \to \mathbb{R}$, the function $f \circ T$ is also integrable and $\int_Y f d\nu = \int_X f \circ T d\mu$.

By picking $h = I_{[\theta,\infty)}$ we obtain that $F(I_{[\theta,\infty)}) = 0$ for every θ. As a result, F is zero on the characteristic function of every interval and also on linear combinations of such characteristic functions.

Recall that simple functions have the form $\phi = \sum_{i=1}^{m} a_i I_{A_i}$ for some $m \in \mathbb{N}$, $a_i \in \mathbb{R}$ and A_i measurable subsets of $\mathbb{R}$. This means that, $F(\phi) = 0$ when ϕ is a simple function whose measurable sets A_i are all intervals. Since we are in $L^1(\mathbb{R}, \nu)$, we can deduce that $F(\phi) = 0$ for every simple ϕ using an ε-argument: Let $A \subseteq \mathbb{R}$ be a Borel set with $\mu(A) > 0$ and let $\varepsilon > 0$. By the outer regularity of μ, there exists some open set $U \supseteq A$ with $\mu(A \setminus U) < \frac{\varepsilon}{2}$. Additionally, U can be written as the disjoint union of open intervals [Arg11, Theorem 6.13], $U = \cup_{k=1}^{\infty} I_k$ and we can pick an n_0 such that $\mu\left(\bigcup_{k=1}^{n_0} I_k\right) \in (\mu(U) - \frac{\varepsilon}{2}, \mu(U) + \frac{\varepsilon}{2})$. So, if we set $J = \cup_{k=1}^{n_0} I_k$ we obtain that $\|I_A - I_J\|_1 \leq \mu(J \triangle A) < \varepsilon$. By the triangle inequality and the fact that $F(I_J) = 0$, the quantity $F(I_A)$ can be made arbitrarily small.

As simple functions are dense in $L^1(\mathbb{R}, \nu)$ [AB06, Theorem 13.8] and F is continuous, we conclude that $F(h) = 0$ for every $h \in L^1(\mathbb{R}, \nu)$. For some fixed $m \in \mathbb{R}$, let $h_1(t) = \sin mt$ and $h_2(t) = \cos mt$. Clearly $h_1, h_2 \in L^1(\mathbb{R}, \nu)$ since both of them are bounded and ν is a finite measure. By applying F to $h_2 + ih_1$, we obtain that

$$\begin{aligned} F(h_1 + ih_2) &= \int_{I_n} \cos(m^\top x) + i \sin(m^\top x) d\mu(x) \\ &= \int_{I_n} e^{im^\top x} d\mu(x) =: \hat{\mu}(m) = 0 \end{aligned}$$

for every $m \in \mathbb{R}$, where $\hat{\mu}(m)$ denotes the Fourier coefficient of the measure μ at m. Since every Fourier coefficient of μ is zero, μ is also zero, thus σ is discriminatory. ■

Theorem 2.2.7: *Every continuous, sigmoidal function σ is discriminatory and the set $\Sigma_n(\sigma)$ is dense in $C(I_n)$.*

Proof. Every continuous, sigmoidal function σ is bounded. Indeed, let $\varepsilon = 1$ and pick $t_0 < t_1 \in \mathbb{R}$ such that $\sigma(t) < 1$ for every $t < t_0$ and $\sigma(t) < 2$ for every $t \geq t_1$. Since $\sigma|_{[t_0,t_1]}$ is continuous, we have that $\sigma(t) \leq M$ for every $t \in [t_0, t_1]$ for some $M > 0$. Therefore, σ is bounded on the whole real line by $\max\{2, M\}$. By the previous theorem, σ is discriminatory and by Theorem 2.2.5, $\Sigma_n(\sigma)$ is dense in $C(I_n)$. ■

2.2.1 *Applications to classification problems*

Let $P_1, \ldots, P_k$ be a partition of I_n and $f : I_n \to \{1, \ldots, k\}$ be the function with the property that $f(x) = j$ if and only $x \in P_j$. In a physical setting, I_n can be viewed as the set consisting of the whole population, whereas each partition P_i represents a subgroup of the original population. The function f assigns each individual to his respective group. We are interested in approximating the function f using

neural networks, however Cybenko's theorem is not directly applicable, as such classification functions are always[8] discontinuous,[9] unless the partition is trivial.

Approximation of classification functions by neural networks is possible, if one is willing to accept an arbitrarily small but predefined probability ε of being wrong. Let λ be the Lebesgue measure, or any other Borel measure on I_n. Using Lusin's Theorem, we can approximate any classification function by a continuous function on a compact subset of I_n that supports most of the measure λ.

Theorem 2.2.8 (Lusin): [AB06, Theorem 12.8] *Let (X,ρ) be a metric space, μ a Borel measure on X and $f : X \to \mathbb{R}$ a measurable function. Then, for every $\varepsilon > 0$, there exists a compact set $K \subset X$ such that $\mu(K^c) < \varepsilon$ and the restriction of f on K is continuous*

Theorem 2.2.9 (Tietze): [Mun13, Theorem 35.1] *Let (X,ρ) be a metric space, $A \subseteq X$ be a closed subset of X and $f : A \to \mathbb{R}$ be a continuous function. Then there exists a continuous function $\tilde{f} : X \to \mathbb{R}$ such that $\tilde{f}(a) = f(a)$ for every $a \in A$.*

Theorem 2.2.10: *Let σ be a continuous sigmoidal function and f be a classification function on some finite partition of I_n. Then for every $\varepsilon > 0$, there exists a function G of the form*

$$G(x) = \sum_{j=1}^{N} a_j \sigma\left(w_j^\top x + \theta_j\right) \tag{2.2.9}$$

and a compact set $K \subseteq I_n$ with $\lambda(K^c) < \varepsilon$, such that $\|G - f\|_\infty < \varepsilon$ on K.

Proof. By Lusin's theorem, there exists some compact set K with the property that $\lambda(K^c) < \varepsilon$ and the restriction of f on K is continuous. By Tietze's extension theorem, there exists some continuous $h : I_n \to \mathbb{R}$ such that $h(x) = f(x)$ for every $x \in K$. By Theorem 2.2.7, there exists some function $G(x)$ of the form (2.2.9), such that $\|G-h\|_\infty < \varepsilon$ on I_n, thus $|G(x)-f(x)| = |G(x)-h(x)| < \varepsilon$ for every $x \in K$. ■

The previous theorem can provide us with a classification rule S that is correct with probability $1-\varepsilon$. This rule is classifying a drawn point x according to the closest integer of the value $G(x)$ and may give a wrong classification only when $x \notin K$.

[8] A topological space (X,τ) is called *connected*, if it cannot be written as the union of two nonempty open sets. For $A \subseteq X$, we define the *boundary* of A as, $\operatorname{bd} A = \overline{A} \setminus \mathring{A}$. In any connected space, every set $\emptyset \neq A \subsetneq X$ has a nonempty boundary. Indeed, if $\operatorname{bd} A = \emptyset$, then $A = \overline{A} = \mathring{A}$ would be a clopen set, and so would $X \setminus A$. Then, $X = A \cup (X \setminus A)$ with both sets being nonempty and open, a contradiction.

[9] Let $P_1, \ldots, P_k$, $k \geq 2$, be a partition of the connected space I_n, and f be a classification function. Pick an $x \in \operatorname{bd} P_1$ and two sequences $(x_n)_n$, $(y_n)_n$ that converge to x, the former belonging in P_1 and the latter in P_1^c. Using the Pigeonhole Principle, we may assume that every element of $(y_n)_n$ belongs to the same partition set, say P_j for some $j \neq 1$. Then $f(x_n) \to 1$, whereas $f(y_n) \to j \neq 1$, so f cannot be continuous.

Corollary 2.2.11: *Let f be a classification function on the partition $P_1, \dots, P_k$ of I_n and suppose that we are sampling according to a Borel probability measure λ on I_n. Then for every $\varepsilon < \frac{1}{2}$, there exists a neural network induced classification rule $S : I_n \to \{1, \dots, k\}$ which is correct at least $(1-\varepsilon)\%$ of the time.*

Proof. Let $\varepsilon < \frac{1}{2}$ and G, K be the function and compact set provided by Theorem 2.2.10 respectively. Suppose that we draw the value $x \in I_n$. Let $S(x)$ denote the index which is closer to $G(X)$, namely $S(x) = \arg\min_{j=1,\dots,k}\{|G(x) - j|\}$. If $x \in K$, then $|G(x) - f(x)| < \varepsilon < \frac{1}{2}$ and $|G(x) - j| > \frac{1}{2} > \varepsilon$ for every $j \neq f(x)$. This implies that $S(x) = f(x)$ for every $x \in K$ and since we are drawing samples from λ, the probability that a drawn value belongs to K, is $\lambda(K) > 1 - \varepsilon$. So the classification rule S is correct at least $(1-\varepsilon)\%$ of the time. ∎

2.2.2 *Measurable activation functions*

In the case where the function f we are trying to approximate is not continuous, but belongs to some $L_p(I_n)$ for $1 \leq p < \infty$, the set $\Sigma_n(\sigma)$ is still dense in L_p under the $\|\cdot\|_p$ norm.

Theorem 2.2.12: *Let $1 \leq p < \infty$, λ be the Lebesgue measure, or any other finite Borel measure, on I_n and σ be a bounded and measurable sigmoidal function. Then the set*

$$\Sigma_n(\sigma) = \left\{ f : I_n \to \mathbb{R} : f(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j) : N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \right\} \tag{2.2.10}$$

is dense in L_p for every $1 \leq p < \infty$.

Proof. Suppose not. Then $\overline{\Sigma_n(\sigma)} \subsetneq L_p$. By the Hahn-Banach theorem, there exists some nonzero linear and bounded functional $F : L_p \to \mathbb{R}$ with $F(f) = 0$ for every $f \in \Sigma_n(\sigma)$. Let q denote the conjugate exponent of p.[10] By the Riesz Representation Theorem for L_p, [RF10, §19.2][11] there exists an $h \in L_q(I_n)$ such that

$$F(f) = \int_{I_n} f(x) h(x) \, d\lambda(x)$$

for every $f \in L_p(I_n)$. Additionally, the fact that F is zero on $\Sigma_n(\sigma)$ implies that

$$\int_{I_n} \sigma(w^\intercal x + \theta) h(x) \, d\mu(x) = 0$$

[10] Two real numbers p, q with $1 \leq p, q \leq \infty$ are said to be *conjugate exponents*, if they satisfy the relation $\frac{1}{p} + \frac{1}{q} = 1$, where for $p = 1$ we have made the convention that $\frac{1}{1} + \frac{1}{\infty} = 1$.

[11] If $1 \leq p, q \leq \infty$ are conjugate exponents, then $L_p(\mu)^*$ is isometrically isomorphic to $L_q(\mu)$ under the operator $T : L_p \to L_q$, defined as $T(g)(f) = \int f g \, d\mu$ for every $f \in L_q$.

for every $w \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$. Let ν be the signed Borel measure induced by h, namely $\nu(A) = \int_A h(x)d\lambda(x)$ for every $A \in B(I_n)$. By Hölder's inequality,

$$\begin{aligned} |\nu(A)| &= \left|\int_A h(x)d\lambda(x)\right| = \left|\int_{I_n} h(x)I_A(x)d\lambda(x)\right| \\ &\leq \int_{I_n} |h(x)|I_A(x)d\lambda(x) \\ &\leq \begin{cases} \|h\|_p \lambda(A)^{1/q} < \infty, & \text{when } p > 1, \\ \|h\|_1 \lambda(A) < \infty, & \text{when } p = 1, \end{cases} \end{aligned}$$

and in either case ν is a finite, Borel signed measure with $\int_{I_n} \sigma(w^\intercal x + \theta)d\nu(x) = 0$ for every $w \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$. As we saw earlier, bounded sigmoidal functions are discriminatory, so $\nu \equiv 0$ and consequently, $F \equiv 0$, a contradiction. ■

Corollary 2.2.13: *Let f be a classification function for the partition $P_1, \ldots, P_k$ of I_n and σ be a bounded and measurable sigmoidal function. For every $\varepsilon > 0$, there exists a function G of the form*

$$G(x) = \sum_{j=1}^{N} a_j \sigma\left(w_j^\intercal x + \theta_j\right) \tag{2.2.11}$$

and a set $D \subseteq I_n$ with $\lambda(D^c) < \varepsilon$ and $\|G - f\|_\infty < \varepsilon$ on D.

Proof. By the previous theorem, there exists a G of the desired form such that $\|G - f\|_p < \varepsilon^{1+1/p}$. By Markov's Inequality,

$$\begin{aligned} P\left(\{x : |G(x) - f(x)| \geq \varepsilon\}\right) &= P\left(\{x : |G(x) - f(x)|^p \geq \varepsilon^p\}\right) \\ &\leq \frac{\|G - f\|_p^p}{\varepsilon^p} < \frac{\left(\varepsilon^{1+1/p}\right)^p}{\varepsilon^p} = \varepsilon, \end{aligned}$$

as promised. ■

As in the continuous case, under the previous assumptions, for every $\varepsilon < \frac{1}{2}$, there always exists a neural network based classification rule that is correct $(1 - \varepsilon)\%$ of the time.

2.2.3 *Stone - Weierstrass Approximation*

As we mentioned in a previous note, in Cybenko's proof one cannot simply apply the Stone-Weierstrass Theorem on the vector space $\Sigma_n(\sigma)$ to deduce that it is dense in $C(I_n)$. Let us see why this is the case and how we can modify the set $\Sigma_n(\sigma)$ for the Stone-Weierstrass Theorem [AB06, p. 352] to be applicable.

Theorem 2.2.14 (Stone-Weierstrass): *Let X be a compact space and A be a subalgebra[12] of $C(X)$ that separates the points[13] of X and contains the constant function* $\mathbf{1}$.[14] *Then A is dense in $C(X)$.*

It is easy to see that $\Sigma_n(\sigma)$ satisfies every condition of the Stone - Weierstrass theorem, except that it may not necessarily be closed under multiplication. As a counterexample, let $n = 1$ and σ be the function

$$\sigma(x) = \begin{cases} 0, & x \leq 0, \\ x, & x \in (0,1), \\ 1, & x \geq 1. \end{cases}$$

We will show that $\sigma(x)^2$ does not belong in $\Sigma_1(s)$. Suppose that it does. Then

$$\sum_{j=1}^{N} a_j \sigma\left(w_j x + \theta_j\right) = \begin{cases} 0, & x \leq 0, \\ x^2, & x \in (0,1), \\ 1, & x \geq 1, \end{cases}$$

for some $N \in \mathbb{N}$, $a_j, \theta_j \in \mathbb{R}$ and $w_j \in \mathbb{R}$. For $j = 1$ and $I_0 = [0,1]$, we can pick a subinterval $I_1 \subseteq I_0$ where $\sigma(w_1 x + \theta_1)$ is either constant, or equal to $w_1 x + \theta_1$. For $j = 2$ and the previously chosen I_1, we can pick a subinterval $I_2 \subseteq I_1$ where $\sigma(w_2 x + \theta_2)$ is either constant, or equal to $w_2 x + \theta_2$. By proceeding inductively, after N steps we end up with an interval I with the property that for every $j = 1, \ldots, N$, the function $\sigma(w_j x + \theta_j)$ is either constant, or equal to $w_j x + \theta_j$ for every $x \in I$. Consequently, if we set $A = \{j : \sigma(w_j x + \theta_j) = 1 \ \forall x \in I\}$ and $B = \{j : \sigma(w_j x + \theta_j) = w_j x + \theta_j \ \forall x \in I\}$, we have that

$$\sum_{j=1}^{N} a_j \sigma\left(w_j x + \theta_j\right) = \sum_{j \in A} a_j + \sum_{j \in B} a_j (w_j x + \theta_j) = x^2, \ \forall x \in I.$$

By the last expression, the polynomials $Q(x) = \sum_{j \in A} a_j + \sum_{j \in B} a_j (w_j x + \theta_j)$ and $P(x) = x^2$ are identical on the open interval I, thus they must be identical everywhere.[15] However $P(x)$ is a polynomial of degree two, whereas $Q(x)$ has a degree of at most one, a contradiction.

[12] Let X be a topological space. The space of continuous functions on X, can be equipped with a multiplication operation "$\cdot$" which is defined pointwise: For every $f, g \in C(X)$ we define $(f \cdot g)(x) = f(x)g(x)$ for every $x \in X$. A subspace A of $C(X)$ is called a *subalgebra*, if it is closed under the multiplication operation. Namely, for every $f, g \in A$, their product $f \cdot g$ also belongs to A.

[13] We say that a set $A \subseteq C(X)$ *separates the points* of X, if for every $x \in X$ there exist $f, g \in A$ such that $f(x) \neq g(x)$.

[14] The constant function $\mathbf{1}$ is defined as $\mathbf{1}(x) = 1$ for every $x \in X$.

[15] If P, Q are polynomials that agree on the open interval I, then $R = P - Q$ is a polynomial for which $R(x) = 0$ for every x in I. Since the only polynomial with infinitely many roots is the zero polynomial, we have that $P = Q$.

Hornik, Stinchcombe and White [HSW89, Theorem 2.1] proved that if we close $\Sigma_n(G)$ under multiplication, we can assure that the resulting set will be dense in $C(I_n)$ for any continuous nonconstant function G, regardless of G being sigmoidal or not:

Definition 2.2.15: For $G : \mathbb{R} \to \mathbb{R}$ measurable, we define $\Sigma\Pi_n(G)$ as

$$\Sigma\Pi_n(G) = \left\{ f : I_n \to \mathbb{R} : f(x) = \sum_{j=1}^{N} a_j \prod_{k=1}^{l_j} G(y_{j,k}^\intercal x + \theta_{j,k}) : \right.$$
$$\left. N, l_j \in \mathbb{N}, a_j, \theta_{j,k} \in \mathbb{R}, y_{j,k} \in \mathbb{R}^n \right\}.$$

Theorem 2.2.16: *If $G : \mathbb{R} \to \mathbb{R}$ is continuous and nonconstant, then $\Sigma\Pi_n(G)$ is dense in $C(I_n)$.*

Proof. Let

$$f(x) = \sum_{j=1}^{N} a_j \prod_{k=1}^{l_j} G(y_{j,k}^\intercal x + \theta_{j,k}), \text{ and } g(x) = \sum_{j=1}^{\tilde{N}} \tilde{a}_j \prod_{k=1}^{\tilde{l}_j} G(\tilde{y}_{j,k}^\intercal x + \tilde{\theta}_{j,k})$$

be two elements in $\Sigma\Pi_n(G)$. Their product can be written as

$$\begin{aligned} f(x)g(x) &= \sum_{j=1}^{N} a_j \prod_{k=1}^{l_j} G(y_{j,k}^\intercal x + \theta_{j,k}) \cdot \sum_{j=1}^{\tilde{N}} \tilde{a}_j \prod_{k=1}^{\tilde{l}_j} G(\tilde{y}_{j,k}^\intercal x + \tilde{\theta}_{j,k}) \\ &= \sum_{j=1}^{N} \sum_{i=1}^{\tilde{N}} a_j \tilde{a}_i \prod_{k=1}^{l_j} G(y_{j,k}^\intercal x + \theta_{j,k}) \cdot \prod_{k=1}^{\tilde{l}_i} G(\tilde{y}_{i,k}^\intercal x + \tilde{\theta}_{i,k}) \\ &= \sum_{j=1}^{N} \sum_{i=1}^{\tilde{N}} a_j \tilde{a}_i \prod_{k=1}^{l_j + \tilde{l}_i} G(z_{i,j,k}^\intercal x + \rho_{i,j,k}), \end{aligned}$$

where $z_{i,j,k} = y_{j,k}$ when $k = 1, \ldots, l_j$ and $z_{i,j,k} = \tilde{y}_{i,k}$ when $k = l_j + i$ for $i = 1, \ldots, \tilde{l}_i$. The ρ's are defined in a similar manner. Clearly fg belongs in $\Sigma\Pi_n(G)$, so $\Sigma\Pi_n(G)$ is an algebra.

Let $x_0 = (x_1, \ldots, x_n) \in I_n$. Since G is nonconstant, there exist $t_1 \neq t_2 \in \mathbb{R}$ such that $G(t_1) \neq G(t_2)$. Pick $w = (1, \ldots, 1) \in \mathbb{R}^n$ and set $b_1 = t_1 - \sum_{i=1}^n x_i$ and $b_2 = t_2 - \sum_{i=1}^n x_i$. Then

$$G(w^\intercal x + b_1) = G(t_1) \neq G(t_2) = G(w^\intercal x + b_2),$$

so the functions $s_1(x) = G(w^\intercal x + b_1)$ and $s_2(x) = G(w^\intercal x + b_2)$ separate x_0. Finally, for $w = (0, \ldots, 0)$ and θ_0 such that $G(\theta_0) \neq 0$, the constant function $G(\theta_0)$ belongs to $\Sigma\Pi_n(G)$, along with its scalar products. In particular, so does $\mathbf{1} = \frac{1}{G(\theta_0)} \cdot G(\theta_0)$. By the Stone-Weierstrass Theorem we conclude that $\Sigma\Pi_n(G)$ is dense in $C(I_n)$. ∎

2.3 *Approximation using nonpolynomial activation functions*

Cybenko's Theorem has been improved upon, both by relaxing its assumptions and by strengthening its conclusions. Chen, Chen and Liu [CCL91] presented an alternative proof of Cybenko's theorem in which the approximation was constructed explicitly, as opposed to the original one which was purely existential. Kurt Hornik [Hor91] relaxed the assumption that σ should be a sigmoidal function and showed that every continuous nonconstant function σ gives rise to a dense set $\Sigma_n(\sigma)$:

Theorem 2.3.1: *If ψ is continuous, bounded and nonconstant, then $\Sigma_n(\psi)$ is dense in $C(X)$ for every compact set $X \subseteq \mathbb{R}^n$.*

The proof relies on the same ideas, however its implementation is not as straightforward.[16] Leshno, Lin, Pinkus and Schocken [LLPS93] improved Hornik's result by showing that, essentially, a function ϕ has the property that $\Sigma_n(\phi)$ is dense in $C(X)$ for every compact $X \subseteq \mathbb{R}^n$, if and only if ϕ is not equal to a polynomial almost everywhere. We devote the rest of the section to their proof.

First let us fix some terminology. Let $\Omega \subseteq \mathbb{R}^n$. The space $L^\infty(\Omega)$ contains all the measurable functions $f : \Omega \to \mathbb{R}$ for which there exists some $M > 0$ such that $|f(x)| \leq M$ for almost every $x \in \Omega$. When $\Omega \subseteq \mathbb{R}^n$ is open, we define $L^\infty_{\text{loc}}(\Omega)$, as the space which contains the measurable functions f for which $f \in L^\infty(K)$ for every compact $K \subseteq \Omega$. In particular, $C(\mathbb{R}^n) \subseteq L^\infty_{\text{loc}}(\mathbb{R}^n)$, as every continuous function is bounded on compact sets.

We will say that a subset $F \subseteq L^\infty_{\text{loc}}(\mathbb{R}^n)$ is dense in $C(\mathbb{R}^n)$, if its closure with respect to the compact topology[17] contains $C(\mathbb{R}^n)$, that is, $C(\mathbb{R}^n) \subseteq \overline{F}$. We use M to denote the set of functions in $L^\infty_{\text{loc}}(\mathbb{R})$ with the property that the closure of the set of their discontinuity points has measure zero,

$$M = \left\{ f \in L^\infty_{\text{loc}}(\mathbb{R}^n) : \lambda(\overline{A_f}) = 0 \right\}. \tag{2.3.1}$$

Any such function $\sigma \in M$ induces a family of functions from $\mathbb{R}^n$ to $\mathbb{R}$, exactly as in the case of Cybenko's theorem:

$$\Sigma_n(\sigma) = \operatorname{span}\left\{ f : \mathbb{R}^n \to \mathbb{R} : f(x) = \sigma\left(w^\intercal x + \theta\right) \text{ for } w \in \mathbb{R}^n, \theta \in \mathbb{R} \right\}.$$

Leshno et. al proved that for a function $\sigma \in M$, the set $\Sigma_n(\sigma)$ is dense in $C(\mathbb{R}^n)$ if and only if σ is not a polynomial. We will use a series of lemmas that simplify the proof significantly:

[16] Using the exact same argument as Cybenko, one suffices to show that *every bounded and nonconstant function is discriminatory* (as opposed to the claim that every sigmoidal function is discriminatory which was proved by Cybenko). In order to prove this, Hornik uses tools from Fourier analysis in the same spirit as Cybenko, but quite more involved.

[17] Due to the fact that some elements of $C(\mathbb{R}^n)$ are unbounded, the supremum norm does not turn $C(\mathbb{R}^n)$ into a normed space. We therefore equip it with a different topology, the topology of compact convergence. According to it, a sequence of functions $(f_k)_k$ in $C(\mathbb{R}^n)$ converges to some $f \in C(\mathbb{R}^n)$ if $(f_k)_k$ converges uniformly to f on every compact subset of $\mathbb{R}^n$.

Lemma 2.3.2: *If $\sigma \in M$ is not a polynomial and $\Sigma_1(\sigma)$ is dense in $C(\mathbb{R})$, then $\Sigma_n(\sigma)$ is dense in $C(\mathbb{R}^n)$ for every $n \in \mathbb{N}$.*

Proof. The subspace

$$V = \operatorname{span}\left\{f_{a,y} : \mathbb{R}^n \to \mathbb{R} : f_{a,y}(x) = g(a^\intercal x) \text{ for some } g \in C(\mathbb{R}), a \in \mathbb{R}^n\right\}$$

is dense in $C(\mathbb{R}^n)$. Indeed, as mentioned in [Pin99], by picking $g(x) = e^x$, we have that the vector space

$$A = \operatorname{span}\left\{h_a : \mathbb{R}^n \to \mathbb{R} : h_a(x) = e^{a^\intercal x} \text{ for some } a \in \mathbb{R}^n\right\}$$

is contained in V. Additionally, it is easy to see that A is a subalgebra of $C(\mathbb{R}^n)$ that satisfies the assumptions of the Stone-Weierstrass theorem, thus $\overline{A} = C(K)$ for every compact $K \subseteq \mathbb{R}^n$, and the same must hold for $\overline{V}$.

Suppose that $\Sigma_1(\sigma)$ is dense in $C(\mathbb{R})$. Let $g \in C(\mathbb{R}^n)$, $K \subseteq \mathbb{R}^n$ be a compact set and $\varepsilon > 0$. Since $\overline{V} = C(K)$, there exists an h of the form $h(x) = \sum_{i=1}^m f_i(a_i^\intercal x)$, such that $\|h - g(x)\|_\infty < \frac{\varepsilon}{2}$ on K, where $f_i \in C(K)$ and $a_i \in \mathbb{R}^n$. The functions $\lambda_i : \mathbb{R}^n \to \mathbb{R}$ defined as $\lambda_i(x) = a_i^\intercal x$ are continuous,[18] so there exist $M_i > 0$ such that $|\lambda_i(x)| \leq M_i \|x\|$ for every $x \in \mathbb{R}^n$ and for $i = 1, \ldots, m$. Since K is compact, there exists some $M > 0$ such that $\|x\| \leq M$ for every $x \in K$, therefore there exist M_i' such that $|\lambda_i(x)| \leq M_i'$ for every $x \in K$ and for $i = 1, \ldots, m$. This implies that there exist intervals $I_i = [a_i, b_i]$ such that $a_i^\intercal x \in I_i$ for every $x \in K$ and $i = 1, \ldots, m$.

By setting $y = a_i^\intercal x$ we observe that $f_i(a_i^\intercal x) = f_i(y)$ for some $y \in I_i$. Since $\Sigma_1(\sigma)$ is dense, there exist $G_i \in \Sigma_1(\sigma)$ such that $\|G_i - f_i\|_\infty < \frac{\varepsilon}{2m}$ on I_i. By the triangle inequality,

$$\begin{aligned}
\left\| g - \sum_{i=1}^m G_i \right\|_\infty &\leq \|g - h\|_\infty + \left\| h - \sum_{i=1}^m G_i \right\|_\infty \\
&\leq \frac{\varepsilon}{2} + \left\| \sum_{i=1}^m f_i(a_i^\intercal x) - \sum_{i=1}^m G_i(x) \right\|_\infty \\
&\leq \frac{\varepsilon}{2} + \sum_{i=1}^m \frac{\varepsilon}{2m} = \varepsilon
\end{aligned}$$

on K. ∎

Lemma 2.3.3: *If $\sigma \in M$ is a nonpolynomial smooth function, then $\Sigma_1(\sigma)$ is dense in $C(\mathbb{R})$.*

Proof. For every w, h and $\theta \in \mathbb{R}$, the function $h = h_{w,h,\theta}$ defined as

$$h(x) = \frac{\sigma((w+h)x + \theta) - \sigma(wx + \theta)}{h}$$

[18] Each λ_i is a linear functional on $\mathbb{R}^n$, and linear functionals on finite dimensional spaces are always continuous.

belongs in $\Sigma_1(\sigma)$, and consequently $x^k\sigma(wx+\theta) = \frac{d^k\sigma(wx+\theta)}{dw^k}$ belongs in $\overline{\Sigma_1(\sigma)}$ for every $k \in \mathbb{N}$. As σ is nonpolynomial, for every k there exists a θ_k such that $\sigma^{(k)}(\theta_k) \neq 0$. By setting $w = 0$ and $\theta = \theta_k$, we obtain that $\lambda_k(x) = x^k\sigma(\theta_k) \in \overline{\Sigma_1(\sigma)}$, which implies that $\overline{\Sigma_1(\sigma)}$ contains all polynomials, thus it is dense in $C(\mathbb{R})$. ■

Theorem 2.3.4 (Dini's Theorem): [AB06, Theorem 2.66] *If a decreasing sequence of continuous functions converges pointwise on a compact set to some continuous function, then the convergence is uniform.*

Lemma 2.3.5: *Suppose that $\sigma \in M$. Then $\sigma * \phi \in \overline{\Sigma_1(\sigma)}$ for every $\phi \in C_C^\infty(\mathbb{R})$, where $C_C^\infty(\mathbb{R})$ denotes the space of smooth real functions with compact support, whereas $(\sigma * \phi)(x) = \int_{\mathbb{R}} \sigma(x-y)\phi(y)\,dy$ is the convolution of the functions σ and ϕ.*

Proof. We will prove the lemma in the simpler case where σ is additionally assumed to be continuous. For the general case, see [LLPS93, Step 4, pp.864-865]. Let ϕ be a smooth function with a compact support K and pick a compact interval $I = [a, b]$ that contains K. We create a sequence of partitions $(P_n)_n$ of I as follows: The first partition $P_1 = \{I\}$ contains only I. The second one contains the two intervals $I_1 = [a, a + \delta/2]$, $I_2 = [a + \delta/2, b]$, where $\delta = b - a$. Similarly, P_3 contains 4 intervals, the first two of which split I_1 in half, whilst the last two split I_2. We proceed inductively.

For every partition $P_n = \{I_1, \ldots, I_{2^n}\}$ we define s_n as:[19]

$$s_n(x) = \sum_{i=1}^{2^n} \max_{y \in \overline{I_i}} \{\sigma(x-y)\phi(y)\}\lambda(I_i).$$

By the way $(P_n)_n$ was constructed, we have that $(s_n)_n$ is a decreasing sequence of continuous[20] functions that converges pointwise to the continuous function $\sigma * \phi$. Additionally, the convergence is uniform on compact sets by Dini's Theorem. Since each s_n belongs to $\Sigma_1(\sigma)$, we conclude that their limit $\sigma * \phi$ belongs to $\overline{\Sigma_1(\sigma)}$. ■

Lemma 2.3.6: *Let $\sigma \in M$ and $f \in \overline{\Sigma_1(\sigma)}$. Then $\overline{\Sigma_1(f)} \subseteq \overline{\Sigma_1(\sigma)}$.*

Proof. Let σ and f as in the statement above and $\varepsilon > 0$. There exists some $G \in \Sigma_1(\sigma)$, which has the form $G(x) = \sum_{i=1}^m a_i\sigma(w_i x + \theta_i)$, with the property that $\sup_{K \subseteq \mathbb{R} \text{ compact}} \|f - G\|_\infty < \varepsilon$. For every $w, \theta \in \mathbb{R}$, the function $\tilde{f}(x) = f(wx+\theta)$ also belongs to $\overline{\Sigma_1(\sigma)}$: If $w = 0$, the claim is trivial. For $w \neq 0$, let $\tilde{G}(x) = G(wx+\theta)$ and $K \subseteq \mathbb{R}$ compact. Then

$$\sup\left\{\left|\tilde{f}(x) - \tilde{G}(x)\right| : x \in K\right\} = \sup\left\{\left|f(wx+\theta) - G(wx+\theta)\right| : x \in K\right\}$$

[19] Each $s_n(x)$ is just the n-th Darboux upper sum of the function $\sigma(x-y)\phi(y)$ and, as is well known, $(s_n(x))_n$ converges to its Darboux integral, which coincides with the Riemann integral $\int \sigma(x-y)\phi(y)\,dy$. Since the convergence holds for every $x \in \mathbb{R}$, we have that $(s_n)_n$ converges pointwise to $\sigma * \phi$.

[20] With a simple sequential argument, it can easily be shown that each $g_i(x) = \max_{y \in \overline{I_i}}\{\sigma(x-y)\phi(y)\}\lambda(I_i)$ is continuous, and so is s_n, being the sum of continuous functions.

$$= \sup\left\{\left|f(y) - G(y)\right| : y \in wK + \theta\right\} < \varepsilon,$$

since the set $wK + \theta$ is compact. The claim we just proved, implies that $\Sigma_1(f) \subseteq \overline{\Sigma_1(\sigma)}$, so $\overline{\Sigma_1(f)} \subseteq \overline{\Sigma_1(\sigma)}$ as promised. ■

Remark 2.3.7: An obvious consequence of the previous lemma is that if $f \in \overline{\Sigma_1(\sigma)}$ is such that $\Sigma_1(f)$ is dense in $C(\mathbb{R})$, then so is $\Sigma_1(\sigma)$.

Lemma 2.3.8: *Let $\sigma \in M$ such that for some $\phi \in C_C^\infty(\mathbb{R})$ their convolution $\sigma * \phi$ is not a polynomial. Then $\Sigma_1(\sigma)$ is dense in $C(\mathbb{R})$.*

Proof. By [AF03, pp. 29-31], $\sigma * \phi \in C^\infty(\mathbb{R})$. Since $\sigma * \phi$ is not a polynomial, it is dense in $C(\mathbb{R})$ by Lemma 2.3.3. Additionally, Lemma 2.3.5 asserts that $\sigma * \phi \in \overline{\Sigma}_1(\sigma)$, thus $\Sigma_1(\sigma)$ is dense in $C(\mathbb{R})$ by Remark 2.3.7. ■

Lemma 2.3.9: *Let $\sigma \in M$ such that for every $\phi \in C_C^\infty(\mathbb{R})$, their convolution $\sigma * \phi$ is a polynomial. Then, there exists an $m \in \mathbb{N}$ such that $\sigma * \phi$ is a polynomial of degree at most m for every $\phi \in C_C^\infty(\mathbb{R})$.*

Proof. Let $C_C^\infty[a,b]$ denote the set of smooth functions supported on the interval $[a,b]$ and suppose that $\sigma * \phi$ is a polynomial for every $\phi \in C_C^\infty([a,b])$. The space $C_C^\infty[a,b]$ is equipped with the following metric:

$$\rho(f,g) = \sum_{i=0}^{\infty} \frac{1}{2^i} \frac{\|f-g\|_i}{1+\|f-g\|_i}, \tag{2.3.2}$$

where $\|f\|_i = \sum_{j=0}^{i} \sup_{x\in[a,b]} \left|f^{(j)}(x)\right|$. It is easy to see that $(C_C^\infty([a,b]), \rho)$ is a Fréchet space.[21]

$$V_k = \left\{\phi \in C_C^\infty([a,b]) : \sigma * \phi \text{ is a polynomial of degree at most } k\right\}.$$

Clearly, V_k is a closed subspace of $C_C^\infty([a,b])$ for every k: Let $(\phi_n)_n$ be a sequence in V_k that converges to some $\phi \in C_C^\infty([a,b])$. Then

$$\begin{aligned}\rho(\sigma*\phi_n, \sigma*\phi) &= \sum_{i=0}^{\infty} \frac{1}{2^i} \frac{\|\sigma*\phi_n - \sigma*\phi\|_i}{1+\|\sigma*\phi_n - \sigma*\phi\|_i} \\ &= \sum_{i=0}^{k} \frac{1}{2^i} \frac{\|\sigma*\phi_n - \sigma*\phi\|_i}{1+\|\sigma*\phi_n - \sigma*\phi\|_i} + \sum_{i=k+1}^{\infty} \frac{1}{2^i} \frac{\|\sigma*\phi_n - \sigma*\phi\|_i}{1+\|\sigma*\phi_n - \sigma*\phi\|_i}.\end{aligned}$$

For $i > k$, we have that

$$\begin{aligned}\|\sigma*\phi_n - \sigma*\phi\|_i = \sum_{j=0}^{k} \sup_{x\in[a,b]} \left|(\sigma*\phi_n)^{(j)}(x) - (\sigma*\phi)^{(j)}(x)\right| + \\ \sum_{j=k+1}^{i} \sup_{x\in[a,b]} \left|(\sigma*\phi)^{(j)}(x)\right|,\end{aligned}$$

[21] A Fréchet space is a completely metrizable locally convex topological vector space (see Appendix A.4).

as $(\sigma * \phi_n)^{(l)} \equiv 0$ for every $l \geq k$. Since $(\rho(\phi_n, \phi))_n$ converges to zero and the sum $\sum_{j=k+1}^{i} \sup_{x\in[a,b]} \left|(\sigma * \phi)^{(j)}(x)\right|$ is independent of n, for every $i \geq k+1$ we have that $\|(\sigma * \phi)^{(i)}\| = 0$, so $\sigma * \phi$ has a degree of at most k.

By our assumption, $C_C^\infty([a,b]) = \bigcup_{k=0}^{\infty} V_k$, and since $(C_C^\infty([a,b]), \rho)$ is complete, by Baire's Category Theorem [Arg11, Theorem 8.18],[22] there exists some $m \in \mathbb{N}$ such that V_m has a nonempty interior. Since V_m is a subspace of a topological vector space with a nonempty interior, we conclude that it has to be equal to the whole space,[23] $V_m = C_C^\infty([a,b])$.

So, the claim has been proven regarding the space $C_C^\infty([a,b])$. The next step is to show that $V_m = C_C^\infty(I)$ for every closed interval I of length equal to $b-a$. This means that the same m works for all intervals of the given length $b-a$. Suppose that $\sigma * \phi$ is a polynomial for every $\phi \in C_c^\infty[a', b']$, where $b' - a' = b - a$. The function $\tilde{\sigma}$ defined as $\tilde{\sigma}(x) = \sigma(x - a + a')$, also belongs to M. Additionally, for every $\phi \in C_C^\infty[a,b]$, the function $\tilde{\phi}$ defined as $\tilde{\phi}(x) = \phi(x - a' + a)$, belongs to $C_C^\infty[a', b']$.

Let $\phi \in C_C^\infty[a,b]$. Then

$$\begin{aligned}(\tilde{\sigma} * \phi)(x) &= \int_a^b \tilde{\sigma}(x-y)\phi(y)\,dy = \int_{a'}^{b'} \tilde{\sigma}(x + a' - a - z)\phi(z - a' + a)\,dz \\ &= \int_{a'}^{b'} \sigma(x-z)\tilde{\phi}(z)\,dz = (\sigma * \tilde{\phi})(x),\end{aligned}$$

where $\tilde{\phi} \in C_C^\infty[a', b']$, and by our assumption the latter is a polynomial. We proved that $\tilde{\sigma} * \phi$ is a polynomial for every $\phi \in C_C^\infty[a,b]$, thus by the previous step, $\tilde{\sigma}$ is a polynomial of degree at most m and so is σ, being just a translation of $\tilde{\sigma}$.

For the final step, suppose that $\sigma * \phi$ is a polynomial for every $\phi \in C_c^\infty[c,d]$, where $d - c$ is arbitrary. We cover $[c,d]$ by finitely many closed intervals $[c,d] = \bigcup_{i=1}^{N} I_i$, each having length $\lambda(I_i) = b - a$. On each I_i we define a $\phi_i \in C_C^\infty(I_i)$ such that $\phi = \sum_{i=1}^{N} \phi_i$. Then $\sigma * \phi = \sum_{i=1}^{N} \sigma * \phi_i$ and by the previous step, each $\sigma * \phi_i$ is a polynomial of degree at most m. ■

Lemma 2.3.10: *If $\sigma \in M$ and $\sigma * \phi$ is a polynomial of degree at most m for every $\phi \in C_C^\infty(\mathbb{R})$, then σ is itself a polynomial of degree at most m.*

Proof. For every $\phi \in C_C^\infty(\mathbb{R})$,

$$0 = (\sigma * \phi)^{(m+1)}(x) = \int \sigma^{(m+1)}(x-y)\phi(y)\,dy = -\int \sigma(x-y)\phi^{(m+1)}(y)\,dy.$$

By [Fri63, pp. 57-59], we conclude that σ is a polynomial of degree at most m. ■

[22] Let (X, ρ) be a complete metric space and $(F_n)_n$ be a sequence of closed sets in X with the property that $X = \cup_{n=1}^{\infty} F_n$. Then there exists some $n_0 \in \mathbb{N}$ such that F_{n_0} has a nonempty interior.

[23] This is a standard fact for normed spaces [Arg04, Proposition 2.3] which remains true for topological vector spaces. The proof is similar, and is based on the fact that every topological vector space contains a neighborhood base at zero, consisting of absorbing sets (see Appendix A.4 for more details).

We can now establish their result:

Theorem 2.3.11: *Let $\sigma \in M$. The set $\Sigma_n(\sigma)$ is dense in $C(\mathbb{R}^n)$ if and only if σ is not a polynomial.*

Proof. If σ is a polynomial of degree k, then for every $w \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$, the function $\sigma(w^\intercal x+\theta)$ is a polynomial of degree at most k, and so is $\sum_{i=1}^m a_i\sigma(w_i^\intercal x+\theta_i)$ for every m, a_i, w_i and θ_i. This implies that $\langle\Sigma(\sigma)\rangle \subseteq P_k$, where P_k is the space of all polynomials of degree at most k. Let $K \subseteq \mathbb{R}^n$ be a compact set. Then

$$\overline{\langle\Sigma(\sigma)_{|K}\rangle} \subseteq \overline{P_{k|K}} = P_{k|K} \subsetneq C(K),$$

since the space $P_{k|K}$ of polynomials in K of degree at most k, is a finite dimensional subspace of $C(K)$ and thus closed [Arg04, Corollary 2.10] in it.

Conversely, suppose that σ is not a polynomial. In view of Lemma 2.3.9, $\sigma * \phi$ cannot be a polynomial for every $\phi \in C_C^\infty(\mathbb{R})$. Therefore, there exists some $\phi \in C_C^\infty(\mathbb{R})$ such that $\sigma * \phi$ is not a polynomial. By Lemma 2.3.8, we have that $\Sigma_1(\sigma)$ is dense in $C(\mathbb{R})$, and by Lemma 2.3.2 we conclude that $\Sigma_n(\sigma)$ is dense in $C(\mathbb{R}^n)$ for every n. ∎

2.4 *A constructive approach for the universal approximation*

As mentioned before, all the universal approximation proofs we presented so far were purely existential. In this section we study an algorithm,[24] proposed by Kwok and Yeung [KY97], that constructs this approximation for functions f that belong to $L_2(X)$ for some compact $X \subseteq \mathbb{R}^n$. The approximation functions used to approach f belong to some set Γ with the property that $\overline{\operatorname{span}\Gamma} = L_2(X)$. For example, Γ could contain functions of the form

$$\left\{f : \mathbb{R}^n \to \mathbb{R} : f(x) = \sigma\left(w^\intercal x + \theta\right) \text{ for } w \in \mathbb{R}^n, \theta \in \mathbb{R}\right\}$$

for some nonpolynomial continuous function, which were proven to share this property during the previous section. However, exact knowledge of the functions contained in Γ is not necessary for their construction to work. In a purely mathematical formulation, we can state the problem as follows:

Problem 2.4.1: Let $X \subseteq \mathbb{R}^d$ be a compact set and let $\Gamma \subseteq L_2(X)$ be a set with the property that its span is dense in $L_2(X)$. Suppose that $f \in L_2(X)$. Construct a sequence $(f_n)_n$ in $\operatorname{span}\Gamma$ such that $f_n \to f$.

The main idea behind the algorithm is easily explained: Suppose that after the n-th step, we have decided that the functions $g_1, \dots, g_n \in \Gamma$ should appear in our representation of f_n, namely $f_n \in \operatorname{span}\{g_1, \dots, g_n\} =: F_n$. In order to determine the coefficients β_i that will appear in the linear combination $f_n = \sum_{i=1}^n \beta_i g_i$, we

[24] In order to follow the proof, the reader should be familiar with some basic results from Hilbert space theory (see Appendix A.3).

take f_n to be equal to the projection of f on F_n. We can compute the coefficients of $f_n = \sum_{i=1}^{n} \lambda_i e_i$ with respect to an orthonormal basis $(e_i)_{i=1}^{n}$ of F_n (Proposition A.3.9), and then rewrite f_n in terms of the g_i's.[25] By taking the projection, we can assure that f_n is the element of F_n which minimizes the distance between f and F_n. In a sense, f_n is the closest representative of F_n to the real function f.

To determine the next element g_{n+1} to be added to the set $\{g_1, \ldots, g_n\}$, we minimize the residual error $\|f - (f_{n-1} + \beta g)\|$ over every $\beta \in \mathbb{R}$ and $g \in \Gamma$. We add the minimizer g_n to our set, $F_{n+1} = \text{span}\{g_1, \ldots, g_n, g_{n+1}\}$ on which we once again take the projection of f, and so on. During each iteration, the set $\{g_1, \ldots, g_n\}$ is enriched, and additionally, the coefficients of each g_i are recalculated so as to minimize the residual error.

Lemma 2.4.2: *For a fixed $g \in \Gamma$, the expression $\|f - (f_{n-1} + \beta g)\|$ attains its minimum m for $\beta^* = \frac{\langle e_{n-1}, g\rangle}{\|g\|^2}$. Additionally, this minimum is equal to $m = \|e_{n-1}\|^2 - 2\frac{\langle e_{n-1}, g\rangle^2}{\|g\|^2}$.*

Proof. The minimum m of the expression $\|f - (f_{n-1} + \beta g)\|$ with respect to β is equal to the distance between the element $f - f_{n-1}$ and the closed subspace generated by g, $F = \langle g \rangle$, and according to the relevant theory (see Proposition A.3.9), m is achieved at the projection $P_F(e_{n-1})$ of $e_{n-1} = f - f_{n-1}$ to F. The space F is spanned by the unit vector $\frac{g}{\|g\|}$, so this projection is equal to

$$P_F(e_{n-1}) = \left\langle e_{n-1}, \frac{g}{\|g\|} \right\rangle \frac{g}{\|g\|} = \frac{\langle e_{n-1}, g\rangle}{\|g\|^2} g,$$

which implies that $\beta^* = \frac{\langle e_{n-1}, g\rangle}{\|g\|^2}$ is the unique minimizer of the expression in question.

For the additional part, the distance $d(e_{n-1}, F)$ between e_{n-1} and F satisfies

$$\begin{aligned} d(e_{n-1}, F)^2 &= \left\| e_{n-1} - \frac{\langle e_{n-1}, g\rangle}{\|g\|^2} g \right\|^2 \\ &= \|e_{n-1}\|^2 + \frac{\langle e_{n-1}, g\rangle^2}{\|g\|^4} \|g\|^2 - 2\frac{\langle e_{n-1}, g\rangle^2}{\|g\|^2} \\ &= \|e_{n-1}\|^2 - \frac{\langle e_{n-1}, g\rangle^2}{\|g\|^2} \end{aligned} \tag{2.4.1}$$

as promised. ∎

By Lemma 2.4.2, given that we have constructed f_{n-1} and the corresponding error $e_{n-1} = f - f_{n-1}$, in order to minimize the expression $\|e_{n-1} - \beta g\|$ over all $\beta \in \mathbb{R}$ and $g \in \Gamma$, we need to minimize (2.4.1), or equivalently, to compute the

[25] The projection of the element f to the subspace $F_n = \langle e_1, \ldots, e_n \rangle$ is equal to $f_n = \sum_{i=1}^{n} \langle f, e_i \rangle e_i$. Since each e_i belongs to F_n and F_n is also spanned by $\{g_1, \ldots, g_n\}$, we can write every e_i as a linear combination of $g_1, \ldots, g_n$, and thus obtain an expression of f involving only the elements g_i, say $f_n = \sum_{i=1}^{n} \beta_i^n g_i$.

$\arg\max_{g\in\Gamma}\left\{\frac{\langle e_{n-1},g\rangle^2}{\|g\|^2}\right\}$. The lack of any analytical structure on Γ excludes, in general, the possibility for this maximum to exist, however we can settle with less: By the Cauchy – Schwarz inequality, the set $\left\{\frac{\langle e_{n-1},g\rangle^2}{\|g\|^2} : g\in\Gamma\right\}$ is bounded by $\|e_{n-1}\|^2$, so its supremum, say M_n, is finite, and picking a g_n such that $\frac{\langle e_{n-1},g_n\rangle^2}{\|g_n\|^2}$ is larger than $M_n/2$ suffices for the proposed algorithm to converge.

The Kwok-Yeung Algorithm	
Step 1	Initialize with $f_0 = 0$ and $e_0 = f - f_0 = f$.
Step 2	Suppose that $f_{n-1} = \sum_{i=1}^{n-1}\beta_i^{n-1}g_i$ has been constructed, with each $g_i\in\Gamma$. Set $e_{n-1} = f - f_{n-1}$, $M_n = \sup_{g\in\Gamma}\left\{\frac{\langle e_{n-1},g\rangle^2}{\|g\|^2}\right\}$ and pick some $g_n\in\Gamma$ such that $\frac{\langle e_{n-1},g_n\rangle^2}{\|g_n\|^2} > \frac{M_n}{2}$.
Step 3	Set $\beta_n^{n-1} = \frac{\langle e_{n-1},g_n\rangle}{\|g_n\|^2}$, $F_n = \operatorname{span}\{g_1,\dots,g_n\}$ and $\tilde{f}_n = f_{n-1} + \beta_n^{n-1}g_n = \sum_{i=1}^{n-1}\beta_i^{n-1}g_i + \beta_n^{n-1}g_n$.
Step 4	Project f to the closed subspace F_n, to obtain an $f_n = \sum_{i=1}^{n}\beta_i^n g_i$.
Step 5	Go to Step 2.

Table 1: The Kwok-Yeung Algorithm [KY97].

Theorem 2.4.3: *The sequence $(f_n)_n$ constructed by the Kwok-Yeung algorithm converges to f.*

Proof. Let $\tilde{f}_n\in F_n$ as defined in Step 3 of the Algorithm. Since f_n is the projection of f on F_n, we have that $\|f - f_n\| \le \|f - \tilde{f}_n\|$. Therefore,

$$\begin{aligned}\|e_{n-1}\|^2 - \|e_n\|^2 &= \|f - f_{n-1}\|^2 - \|f - f_n\|^2\\ &\ge \|f - f_{n-1}\|^2 - \|f - \tilde{f}_n\|^2\\ &= \|e_{n-1}\|^2 - \|\tilde{e}_n\|^2,\end{aligned} \tag{2.4.2}$$

where $e_{n-1} = f - \sum_{i=1}^{n-1}\beta_i^{n-1}g_i$ and $\tilde{e}_n = f - \sum_{i=1}^{n-1}\beta_i^{n-1}g_i - \beta_n^{n-1}g_n$. Since $\tilde{e}_n \perp g_n$ and $\tilde{e}_n = e_{n-1} - \beta_n^{n-1}g_{n-1}$, we have that $\tilde{e}_n \perp \tilde{e}_n - e_{n-1}$, so by the Pythagorean Theorem,

$$\|e_{n-1}\|^2 = \|\tilde{e}_n + e_{n-1} - \tilde{e}_n\|^2 = \|\tilde{e}_n\|^2 + \|e_{n-1} - \tilde{e}_n\|^2,$$

which implies that

$$\|e_{n-1}\|^2 - \|\tilde{e}_n\|^2 = \|\tilde{e}_n - e_{n-1}\|^2 = \|\beta_n^{n-1}g_n\|^2 = \frac{\langle e_{n-1},g_n\rangle^2}{\|g_n\|^2}.$$

This, combined with (2.4.2), yields that

$$\|e_{n-1}\|^2 - \|e_n\|^2 \ge \frac{\langle e_{n-1},g_n\rangle^2}{\|g_n\|^2} \ge 0 \tag{2.4.3}$$

for every $n \in \mathbb{N}$. Notice here that if $\frac{\langle e_{n-1}, g_n\rangle^2}{\|g_n\|^2} = 0$ for some n, this would imply that $\langle e_{n-1}, g\rangle = 0$ for every $g \in \Gamma$, which would in turn force e_{n-1} to be equal to zero.[26] In this case, $f_{n-1} = f$ and the algorithm terminates having found the exact form of f. So, lets assume that (2.4.3) is strictly greater than zero for every $n \in \mathbb{N}$. Then the real sequence $(\|e_n\|)_n$ is strictly decreasing, thus converges to its infimum. In particular, for every $\varepsilon > 0$ there exists some $n_0 \in \mathbb{N}$ such that $\|e_n\|^2 - \|e_m\|^2 < \varepsilon$ for every $m > n > n_0$.

By Step 4 of the algorithm, each e_n is orthogonal to $\{g_1, \ldots, g_n\}$, and for every $m > n > n_0$, $e_m - e_n \in \operatorname{span}\{g_1, \ldots, g_m\}$, so $e_m \perp e_m - e_m$. Again by the Pythagorean Theorem we obtain that $\|e_n - e_m\|^2 = \|e_n\|^2 - \|e_m\|^2 < \varepsilon$. This implies that $(e_n)_n$ is a Cauchy sequence in the Banach space $L_2(X)$ and so it converges to some $e \in L_2(X)$.

By taking limits in (2.4.3), we obtain that $\lim_{n\to\infty} \frac{\langle e_{n-1}, g_n\rangle^2}{\|g_n\|^2} = 0$. Furthermore, for every $g \in \Gamma$, we have that $\frac{\langle e_{n-1}, g\rangle^2}{\|g\|^2} \leq \frac{\langle e_{n-1}, g_n\rangle^2}{\|g_n\|^2}$ for every n, so by the continuity of the inner product,

$$\frac{\langle e, g\rangle^2}{\|g\|^2} = \lim_{n\to\infty} \frac{\langle e_{n-1}, g\rangle^2}{\|g\|^2} \leq \lim_{n\to\infty} \frac{\langle e_{n-1}, g_n\rangle^2}{\|g_n\|^2} = 0.$$

Thus, $\langle e, g\rangle = 0$ for every $g \in \Gamma$, and as $\overline{\operatorname{span}\Gamma} = L_2(X)$, we conclude that $e = 0$. ■

Remark 2.4.4: The previous algorithm requires the knowledge of f in each computation involving e_n, which in real life problems is not possible. Kwok and Yeung propose how to estimate the e_n's using appropriate consistent estimators based on information available from the training data set [KY97, p. 1134].

2.5 *The Kolmogorov-Arnold approximation*

In their efforts to answer Hilbert's famous 13th problem, Andrey Kolmogorov and Vladimir Arnold published a series of articles in the late '50s, which resulted in one of the most surprising and elegant results of modern mathematics. Hilbert's conjecture was that the roots of the equation $x^7 + ax^3 + bx^2 + cx + 1 = 0$, seen as a function of the parameters a, b, c, could not be written as a superposition of functions of two variables. Arnold [Arn57] disproved this conjecture, but Kolmogorov [Kol57] went one step further, showing that any multivariate function $f : [0,1]^n \to \mathbb{R}$ can be written as a superposition of functions of just a single variable.

Theorem 2.5.1 (Kolmogorov-Arnold): [Kol57] *There exist universal constants $\lambda_1, \ldots, \lambda_d$ with $\sum_{j=1}^{d} \lambda_j \leq 1$, and continuous, strictly increasing universal functions*

[26] By the linearity of the inner product, $\langle e_{n-1}, g\rangle = 0$ for every $g \in \operatorname{span}\Gamma$, and by its continuity, $\langle e_{n-1}, g\rangle = 0$ for every $g \in \overline{\operatorname{span}\Gamma} = L_2$. As a result, $\langle e_{n-1}, e_{n-1}\rangle = 0$, which implies that $e_{n-1} = 0$.

$\phi_1, \ldots, \phi_{2d+1}$ which map $[0,1]$ to itself, such that every function $f \in C[0,1]^d$ can be represented as

$$f(x_1, \ldots, x_d) = \sum_{i=1}^{2d+1} g\left(\sum_{j=1}^{d} \lambda_j \phi_i(x_j)\right), \tag{2.5.1}$$

where $g \in C[0,1]$ is a continuous real function that depends on f.

In 1987, two years before Cybenko's theorem, Robert Hecht-Nielsen published a short note [Hec87] in which he pointed out the resemblance between the expression in (2.5.1) and the one appearing in the neural network functions

$$\sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j), \tag{2.5.2}$$

suggesting the possibility of using the Kolmogorov-Arnold Theorem to obtain representation results in this framework. However, the functions ϕ_i appearing in (2.5.1) were highly non-smooth, so an exact representation of this form using sigmoidal functions in the place of the ϕ_i's seemed to be excluded. But since the goal was to approximate functions, rather than to actually represent them, the functions ϕ_i could instead serve as targets that one could try to approach, using sigmoidal ones.

This idea was formulated and implemented four years later by Věra Kůrková ([Kur91], [Kur92]), who also provided upper bounds for the number of nodes needed. This number was not fixed, but depended on the level of accuracy and on some smoothness properties of the target function f.

A common theme in all theorems we presented so far, is that in the expression $\sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j)$, the number of nodes N depended on the choice of the activation function σ and the desired accuracy level ε, and in general, it could get arbitrarily large. Perhaps at the other side of the spectrum, Vitaly Maiorov and Allan Pinkus (1999) [MP99] constructed a specific activation function σ which has the property that any continuous function $f : [0,1]^d \to \mathbb{R}$ can be approached by a two layer neural network which uses at most $3d$ and $6d+3$ nodes at each layer respectively. Their work was also based on the Kolmogorov-Arnold representation theorem.

Theorem 2.5.2 (Maiorov-Pinkus): [MP99] *There exists a smooth sigmoidal activation function σ such that for any $d \in \mathbb{N}$, any compact $K \subseteq \mathbb{R}^d$, any $f \in C(K)$ and any $\varepsilon > 0$, there exist real constants d_i, c_{ij}, θ_{ij}, γ_i and vectors $w_{ij} \in \mathbb{R}^d$, such that*

$$\left| f(x) - \sum_{i=1}^{6d+3} d_i \sigma\left(\sum_{j=1}^{3d} c_{ij} \sigma\left(w_{ij}^\intercal x + \theta_{ij}\right) + \gamma_i\right)\right| < \varepsilon \tag{2.5.3}$$

for every $x \in K$.

Proof. We will prove the result for $K = [-1,1]$. Recall that $C[-1,1]$ is a separable space [AB06, Lemma 3.99] and that one can easily find a countable dense set in

$C[-1,1]$ consisting of smooth functions. Indeed, the set consisting of the polynomials with rational coefficients

$$\mathbb{Q}[x] = \left\{ \sum_{i=0}^{n} q_i x^i : n \in \mathbb{N}, q_i \in \mathbb{Q} \right\},$$

is countable and dense in $C[-1,1]$ by the Weierstrass Theorem. So, let $(u_k)_k$ be a countable and dense subset of $C[-1,1]$ consisting of smooth functions. We begin by showing that there exists some strictly increasing and smooth sigmoidal function ϕ, such that for every $g \in C[-1,1]$ and $\varepsilon > 0$, there exists an $m \in \mathbb{N}$ and a, b, $c \in \mathbb{R}$ with

$$\left| g(t) - \left(a\phi(t-7) + b\phi(t-3) + c\phi(t+4m+1) \right) \right| < \varepsilon \quad \forall t \in [-1,1]. \tag{2.5.4}$$

Consider a strictly increasing and smooth sigmoidal function h, for example, one can take $h(x) = \frac{1}{1+e^{-x}}$. On the intervals of the form $I_k = [4k, 4k+2]$ for $k = 0, 1, \ldots$, we define ϕ as

$$\phi(t+4k+1) = b_k + c_k t + d_k u_k(t) \text{ for } t \in [-1,1],$$

where b_k, c_k and d_k are coefficients, chosen so that $\phi(4k) = h(4k)$ and $0 < \phi'(t) \leq h'(t)$ for all $t \in I_k$. Such coefficients always exist: By the continuity of h' and u_k' on the compact set I_k, there exists some $m_k < M_k$ with $m_k \leq h'(t) \leq M_k$ for every $t \in I_k$, and similarly some $a_k < A_k$ with $a_k \leq u_k'(t) \leq A_k$. So our two requirements can be met, if both

$$b_k - c_k + d_k u_k(-1) = h(4k) \text{ and } c_k + d_k A_k \leq m_k$$

can be solved simultaneously, which is clearly the case. On each the intervals $J_1 = [-4,-2]$ and $J_2 = [-8,-6]$, ϕ is an affine function that satisfies the same two properties. Additionally, if $\phi(t-3)$ and $\phi(t-7)$ denote the restrictions of ϕ on these two intervals for $t \in [-1,1]$, without loss of generality, we may assume that they are linearly independent (if they are not, just scale one of them accordingly).

On the rest of the real line, ϕ is extended in a way that it remains smooth, strictly monotone and, additionally, $\lim_{t \to -\infty} \phi(t) = 0$. Notice that for every k, there exist $a_k, b_k, c_k \in \mathbb{R}$ such that

$$u_k(t) = (a_k \phi(t-7) + b_k \phi(t-3) + c_k \phi(t+4m+1) \tag{2.5.5}$$

for $t \in [-1,1]$, so our claim has been proven.

Now we return to our proof. Let $f \in C[0,1]$ and $\varepsilon > 0$. By the Kolmogorov-Arnold theorem, f can be represented as

$$f(x_1, \ldots, x_d) = \sum_{i=1}^{2d+1} g\left(\sum_{j=1}^{d} \lambda_j \phi_i(x_j) \right). \tag{2.5.6}$$

By the density of $(u_k)_k$, there exists some u_k with $\|g-u_k\|_\infty < \frac{\varepsilon}{2(2d+1)}$. Applying (2.5.5) for this specific u_k, provides $a_1, a_2, a_3, m_1, m_2, m_3 \in \mathbb{R}$, such that

$$\left|g(t) - (a_1\phi(t+m_1) + a_2\phi(t+m_2) + a_3\phi(t+m_3)\right| < \frac{\varepsilon}{2(2d+1)} \tag{2.5.7}$$

for all $t \in [-1,1]$. Then, (2.5.6) and (2.5.7), combined with the triangle inequality, yield that for every $x \in [0,1]^d$,

$$\begin{aligned}
&\left|f(x) - \sum_{i=1}^{2d+1}\left[a_1\phi\left(\sum_{j=1}^{d}\lambda_j\phi_i(x_j)+m_1\right)+\right.\right.\\
&\qquad\left.\left. a_2\phi\left(\sum_{j=1}^{d}\lambda_j\phi_i(x_j)+m_2\right)+\ a_3\phi\left(\sum_{j=1}^{d}\lambda_j\phi_i(x_j)+m_3\right)\right]\right| =\\
&= \left|\sum_{i=1}^{2d+1} g\left(\sum_{j=1}^{d}\lambda_j\phi_i(x_j)\right) - \sum_{i=1}^{2d+1}\left[a_1\phi\left(\sum_{j=1}^{d}\lambda_j\phi_i(x_j)+m_1\right)+\right.\right.\\
&\qquad\left.\left. a_2\phi\left(\sum_{j=1}^{d}\lambda_j\phi_i(x_j)+m_2\right)+a_3\phi\left(\sum_{j=1}^{d}\lambda_j\phi_i(x_j)+m_3\right)\right]\right| =\\
&= \left|\sum_{i=1}^{2d+1}\left[g(t) - \left(a_1\phi(t+m_1)+a_2\phi(t+m_2)+a_3\phi(t+m_3)\right)\right]\right| \leq\\
&\qquad \leq \sum_{i=1}^{2d+1}\frac{\varepsilon}{2(2d+1)} = \frac{\varepsilon}{2}.
\end{aligned}$$

Therefore,

$$\left|f(x) - \sum_{i=1}^{6d+3} d_i\phi\left(\sum_{j=1}^{d}\lambda_j\phi_i(x_j)+\gamma_i\right)\right| < \frac{\varepsilon}{2} \quad \text{for all } x\in[0,1]^d, \tag{2.5.8}$$

where each of the functions $(\phi_i)_{i=1}^{6d+3}$ is equal to one of the original functions $\{\phi_1,\dots,\phi_d\}$. In turn, for each $\delta>0$ and $i=1,\dots,6d+3$, there exists $b_{i1}, b_{i2}, b_{i3} \in \mathbb{R}$ and $r_{i1}, r_{i2}, r_{i3} \in \mathbb{N}$, such that

$$\left|\phi_i(x_j) - \left((b_{i1}\phi(x_j+r_{i1}) + b_{i2}\phi(x_j+r_{i2}) + b_{i3}\phi(x_j+r_{i3})\right)\right| < \delta$$

for every $x_j \in [0,1]$. Multiplying by λ_i and summing with respect to $i=1,\dots,d$, yields that

$$\left|\sum_{i=1}^{d}\lambda_j\phi_i(x_j) - \sum_{i=1}^{3d} c_{ij}\left(\phi(e_j^\intercal x+\theta_{ij})\right)\right| < \delta \quad \text{for all } x\in[0,1]^d. \tag{2.5.9}$$

As ϕ is continuous, its restriction on any compact set is a uniformly continuous function [Arg11, Theorem 9.27], so we can pick a δ sufficiently small so that for all d_j,

$$\left|\phi\left(\sum_{j=1}^{d}\lambda_j\phi_i(x_j)\right) - \phi\left(\sum_{j=1}^{3d} c_{ij}\left(\phi(e_j^\intercal x+\theta_{ij})\right)\right)\right| < \frac{\varepsilon}{2d_j(6\delta+3)} \quad \text{for all } x\in[0,1]^d$$

and thus,

$$\left|\sum_{i=1}^{6d+3} d_i\phi\left(\sum_{j=1}^{d}\lambda_i\phi_i(x_j)+\gamma_i\right) - \sum_{i=1}^{6d+3} d_i\phi\left(\sum_{j=1}^{3d} c_{ij}\left(\phi(e_j^\intercal x+\theta_{ij})\right)+\gamma_i\right)\right| < \frac{\varepsilon}{2} \tag{2.5.10}$$

for all $x \in [0,1]^d$. The result then follows from (2.5.8), (2.5.10) and the triangle inequality. ■

References

[AF03] R. Adams, J. Fournier, *Sobolev Spaces*, Academic Press, 2rd ed., 2003. ISBN: 9780120441433 Cited on p. 34

[AB06] C. Aliprantis, K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, 3rd ed., 2006. DOI: 10.1007/978-3-662-03004-2 Cited on p. 21, 22, 24, 25, 26, 28, 33, 40, 90, 126, 131, 133

[Arg04] S. Argyros, *Lecture Notes in Functional Analysis* (in greek), Athens, 2004. https://semfe.gr/files/users/795/synarthsiakh_analysh_I_biblio.pdf Cited on p. 22, 35, 36, 54, 121, 122, 123, 125

[Arg11] S. Argyros, *Lecture Notes in Real Analysis* (in greek), Athens, 2011. https://semfe.gr/files/users/795/pragmatiki_book.pdf Cited on p. 25, 35, 42, 132

[Arn57] V. Arnold, On functions of three variables, *Dokl. Akad. Nauk SSSR*, **114**, pp. 679-681, 1957. DOI: 10.1007/978-3-642-01742-1_2 Cited on p. 5, 39

[BT96] D. Bertsekas, J. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996. ISBN: 1-886529-10-8 Cited on p. 11, 19, 54, 75, 80, 82

[CCL91] T. Chen, H. Chen, R. Liu, A Constructive Proof and An Extension of Cybenko's Approximation Theorem, in: *Computing Science and Statistics, Proceedings of the 22nd Symposium on the Interface*, Springer, Berlin, pp. 163–168, 1991. DOI: 10.1007/978-1-4612-2856-1_21 Cited on p. 4, 31

[Cyb89] G. Cybenko, Approximation by Superposition of a Sigmoidal Function, *Mathematics of Control, Signal, and Systems*, **2**, 1989, pp. 303-314. DOI: 10.1007/BF02551274 Cited on p. 2, 21

[Fri63] A. Friedman, *Generalized Functions and Partial Differential Equations,* Prentice Hall, 1963. Cited on p. 35

[Hec87] R. Hecht-Nielsen, Kolmogorov's mapping neural network existence theorem, in: *Proceedings of the International Conference on Neural Networks*, IEEE Press, 1987. ID: 118526925 Cited on p. 6, 40

[HSW89] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Networks,* **2** (5), pp. 359-366, 1989. DOI: 10.1016/0893-6080(89)90020-8 Cited on p. 30

[Hor91] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks,* **4** (2), pp. 251-257, 1991. DOI: 10.1016/0893-6080(91)90009-T Cited on p. 31

[Kol57] A. Kolmogorov, On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition, *Dokl. Akad. Nauk SSSR,* **114** (5), pp. 953–956, 1957. MR111809 Cited on p. 5, 39

[Kur91] V. Kurkova, Kolmogorov's theorem is relevant, *Neural Computation,* **3** (4), pp. 617-622, 1991. DOI: 10.1162/neco.1991.3.4.617 Cited on p. 6, 40

[Kur92] V. Kurkova, Kolmogorov's theorem and multilayer neural networks, *Neural Networks,* **5** (3), pp. 501-506 , 1992. DOI: 10.1016/0893-6080(92)90012-8 Cited on p. 6, 40

[KY97] T.-Y. Kwok, D.-Y. Yeung, Objective functions for training new hidden units in constructive neural networks, *IEEE Trans Neural Netw.,* **8** (5), pp. 1131-1148, 1997. DOI: 10.1109/72.623214 Cited on p. xiv, 4, 36, 38, 39

[LLPS93] M. Leshno, V. Lin, A. Pinkus, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, *Neural Networks,* **6** (6), pp. 861-867, 1993. DOI: 10.1016/S0893-6080(05)80131-5 Cited on p. 3, 31, 33

[MP99] V. Maiorov, A. Pinkus, Lower Bounds for Approximation by MLP Neural Networks, *Neurocomputing,* **25**, pp. 81-91, 1999. DOI: 10.1.1.6.5161 Cited on p. 6, 7, 40

[Mun13] J. Munkres, *Topology,* Pearson New International Ed., 2013. Cited on p. 26

[Pin99] A. Pinkus, Approximation theory of the MLP model in neural networks, *Acta Numerica,* **8**, pp. 143-195, 1999. DOI: 10.1017/S0962492900002919 Cited on p. 32

[RF10] H. Royden, P. Fitzpatrick, *Real Analysis,* Prentice Hall, 4th Ed., 2010. Cited on p. 27

Part II

REINFORCEMENT LEARNING

3

Stochastic Approximation

A common way of solving optimization problems, is through iteration methods. Stochastic iterative algorithms are extensions of these methods in the presence of randomness. In Section 3.1, we present the classic Robbins-Monro algorithm, which is one of the first algorithms developed in this setting. In Section 3.2 we study the basic tools on which many modern stochastic approximation methods are based. The main goal is to prove the convergence of the Q-Learning algorithm, which is presented in Section 3.3.

3.1 *The Robbins-Monro algorithm*

Suppose that for each $x \in \mathbb{R}$ we have a random variable $Y = Y(x)$ with distribution $P[Y(x) \leq y] = H(y \mid x)$, and let $M(x)$ denote the expectation of Y given x, namely $M(x)$ is given by $M(x) = \mathbb{E}[Y \mid X = x] = \int_{\mathbb{R}} y dH(y \mid x)$. Such an hierarchy is common in applications. For example, one could be interested in calculating the effect of a treatment on a patient. If Y is the treatment duration and X is the quantity of the prescribed drug, one may have that given $X = x$, the distribution of Y is a known distribution, having a parameter that depends on x. In some cases, one could choose the family $(Y(x))_x$ carefully, so that $M(x)$ could be computed explicitly, but in general this may not be possible.

In our setting, we presume that we do not know the exact forms of $M(x)$ or $H(y \mid x)$, although we have a method to sample from the distribution $H(y \mid x)$ for every x, and we want to solve the equation $M(\theta) = a$ for some fixed a. A practical reason for solving this problem is that, assuming $M(x)$ is a nondecreasing function, by solving the equation $M(\theta) = a$ we can find a threshold value θ for X, for which the expected response is at least a. This is very useful in applications were we are looking for conditions on X so that a certain expected level of efficiency can be achieved for Y.

In this section we present a solution of this problem using the Robbins-Monro algorithm [RM51]. According to it, a sequence $(x_n)_n$ is constructed inductively based on the formula

$$x_{n+1} = x_n + a_n(a - y_n), \tag{3.1.1}$$

where y_n is a random variable with distribution $H(y \mid x_n)$ and $(a_n)_n$ is a predetermined sequence of nonnegative numbers. Under certain assumptions on the distribution of each $Y(x)$, and by picking $(a_n)_n$ appropriately, Herbert Robbins and Sutton Monro showed that x_n converges in probability to θ.

The number x_n in (3.1.1) is the current estimate of θ. This value is being corrected by $a_n(a-y_n)$ to produce the next estimate x_{n+1}. The quantity $a-y_n$ is the difference between the drawn value y_n and the target value a. Each a_n acts as a weight that adjusts the impact of $a-y_n$ on the output x_{n+1}.[1] The role of the sequence $(a_n)_n$ is subtle, as it serves two different, and rather conflicting, purposes. It needs to be large enough to allow the correction term $a-y_n$ to express itself, but also small enough to not waste the progress made thus far by x_n. Additionally, very small values of $(a_n)_n$ can make the convergence of $(x_n)_n$ easier, but may also confine the algorithm's range to some interval that does not contain θ. On the other hand, large values of $(a_n)_n$, expand the search of θ to a larger domain, but may exhibit poor convergence properties.

The compromise between these two effects is achieved by picking a sequence $(a_n)_n$ that converges to zero, but at a slow pace. In the Robbins - Monro algorithm, as well as in a large variety of similar stochastic approximation algorithms, $(a_n)_n$ is chosen to be a square summable, but not an absolutely summable sequence: $\sum_{n=1}^{\infty} a_n^2 < \infty$, but $\sum_{n=1}^{\infty} a_n = \infty$.

We will present the proof, based on a series of lemmas. Firstly, in the Robbins-Monro paper [RM51, p. 403, relation (26)], the sequence $(a_n)_n$ is chosen so that the series $\sum_{n=1}^{\infty} \frac{a_n}{S_n}$ diverges, where S_n denotes the n-th partial sum of the sequence $(a_k)_k$. This condition is equivalent to the divergence of $\sum_{n=1}^{\infty} a_n$:

Lemma 3.1.1: *Let $(a_n)_n$ be a sequence of nonnegative real numbers and $S_n = a_1 + \ldots + a_n$ be their n-th partial sum. Then $\sum_{n=1}^{\infty} a_n = \infty$ if and only if $\sum_{n=2}^{\infty} \frac{a_n}{S_{n-1}} = \infty$.*

Proof. Suppose that $\sum_{n=2}^{\infty} \frac{a_n}{S_{n-1}} = \infty$, yet the sum $\sum_{n=1}^{\infty} a_n$ does not diverge. As all the terms of $(a_n)_n$ are nonnegative, the sequence of the partial sums $(S_n)_n$ is increasing, thus converges to its supremum, say $\sum_{n=1}^{\infty} a_n = R < \infty$. We pick an $n_0 \in \mathbb{N}$ such that $\frac{R}{2} \le S_n \le \frac{3R}{2}$ for every $n \ge n_0$. This implies that

$$\begin{aligned}\sum_{n=2}^{\infty} \frac{a_n}{S_{n-1}} &= \sum_{n=2}^{n_0} \frac{a_n}{S_{n-1}} + \sum_{n=n_0+1}^{\infty} \frac{a_n}{S_{n-1}} \\ &= \Lambda_0 + \sum_{n=n_0+1}^{\infty} \frac{a_n}{S_{n-1}} \\ &\le \Lambda_0 + \sum_{n=n_0+1}^{\infty} \frac{a_n}{R/2} \le \Lambda_0 + 2 < \infty,\end{aligned}$$

a contradiction.

For the converse, suppose that $\sum_{n=1}^{\infty} a_n = \infty$ and let $p \in \mathbb{N}$. Then

$$\frac{a_n}{S_{n-1}} + \ldots + \frac{a_{n+p}}{S_{n+p-1}} \ge \frac{1}{S_{n+p-1}}(a_n + \ldots + a_{n+p}) = \frac{S_{n+p} - S_{n-1}}{S_{n+p-1}},$$

[1] In the relation (3.1.1), x_{n+1} can be rewritten as a convex combination of the elements x_n and $a-y_n+x_n$ when $a_n \in [0,1]$: Specifically, $x_{n+1} = (1-a_n)x_n + a_n(a-y_n+x_n)$. These two elements, x_n and $a-y_n+x_n$, represent the two extreme proposed values for x_{n+1}. The former suggests that x_{n+1} should completely ignore the correction term $a-y_n$, whereas the latter suggests that x_{n+1} should fully embrace it. The value of a_n determines what weight will be given to each of these two suggestions.

with

$$\lim_{p\to\infty}\frac{S_{n+p}-S_{n-1}}{S_{n+p-1}}=\lim_{p\to\infty}\frac{S_{n+p}}{S_{n+p-1}}-\lim_{p\to\infty}\frac{S_{n-1}}{S_{n+p-1}}\geq 1, \tag{3.1.2}$$

since $\frac{S_{n+p}}{S_{n+p-1}}\geq 1$ for all p and $\lim_{p\to\infty}S_{n+p-1}=\infty$. Inequality (3.1.2) suggests that the sequence of the partial sums of the series $\sum_{n=2}^{\infty}\frac{a_n}{S_{n-1}}$ is not a Cauchy sequence, and since it is increasing, it must be divergent. ■

Lemma 3.1.2: *Suppose that there exists some $C>0$ such that for every $x\in\mathbb{R}$, $P[|Y(x)|\leq C]=1$. Additionally, $M(x)=\mathbb{E}[Y\mid X=x]$ satisfies the property that $M(x)\leq a$ for $x<\theta$ and $M(x)\geq a$ for $x>\theta$. Suppose that $(a_n)_n$ is a square summable sequence and let*

$$d_n=\mathbb{E}[(x_n-\theta)(M(x_n)-a)] \text{ and}$$
$$e_n=\mathbb{E}\left[\int_{-\infty}^{\infty}(y-a)^2dH(y\mid x_n)\right].$$

If there exists some sequence $(k_n)_n$ of nonnegative numbers such that $d_n\geq k_nb_n$ and $\sum_{n=1}^{\infty}a_nk_n=\infty$, then the sequence $(x_n)_n$ constructed by Robbins-Monro algorithm,

$$x_{n+1}=x_n+a_n(a-y_n), \tag{3.1.3}$$

converges in probability to θ.[2]

Proof. Let $b_n=\mathbb{E}[(x_n-\theta)^2]$. We will show that $b_n\to 0$. By substituting (3.1.3) into $\mathbb{E}[(x_n-\theta)^2]$, and conditioning on x_n,[3] we obtain that $b_{n+1}=\mathbb{E}[(x_{n+1}-\theta)^2]=\mathbb{E}[\mathbb{E}[(x_{n+1}-\theta)^2\mid x_n]]$, with

$$\begin{aligned}\mathbb{E}[(x_{n+1}-\theta)^2\mid x_n]&=\mathbb{E}[(x_n-\theta+a_n(a-y_n))^2\mid x_n]\\&=\mathbb{E}[(x_n-\theta)^2+a_n^2(a-y_n)^2-\\&\qquad-2a_n(x_n-\theta)(y_n-a)\mid x_n]\\&=(x_n-\theta)^2+a_n^2\int(a-y)^2dH(y\mid x_n)\\&\qquad-2a_n(x_n-\theta)(M(x_n)-a).\end{aligned}$$

Thus,

$$\begin{aligned}b_{n+1}&=\mathbb{E}[\mathbb{E}[(x_{n+1}-\theta)^2\mid x_n]]\\&=\mathbb{E}[(x_n-\theta)^2]+a_n^2\,\mathbb{E}\left[\int(a-y)^2dH(y\mid x_n)\right]\end{aligned}$$

[2] A sequence of random variables (X_n) is said to converge in probability to X, denoted as $X_n\xrightarrow{P}X$, if for every $\varepsilon>0$, $\lim_{n\to\infty}P[|X_n-X|\geq\varepsilon]=0$.

[3] By the Law of total expectation, if X and Y are random variables, then $\mathbb{E}[X]=\mathbb{E}[\mathbb{E}[X\mid Y]]$ as long as $\mathbb{E}[X]$ exists.

$$
\begin{aligned}
&\quad - 2a_n \mathbb{E}[(x_n - \theta)(a - M(x_n))] \\
&= b_n + a_n^2 \mathbb{E}\left[\int_{-\infty}^{\infty} (y-a)^2 dH(y \mid x_n)\right] \\
&\quad - 2a_n \mathbb{E}[(x_n - \theta)(M(x_n) - a)] \\
&= b_n + a_n^2 e_n - 2a_n d_n,
\end{aligned}
$$

and

$$
b_{n+1} - b_n = a_n^2 e_n - 2a_n d_n. \tag{3.1.4}
$$

Since $x_n - \theta$ and $M(x_n) - a$ have the same sign, d_n is always nonnegative. Additionally, by our assumptions on Y, $-C \leq Y(x) \leq C$ for every x a.e., which implies that $-C - |a| \leq |Y - a| \leq C + |a|$ a.e., so $|Y - a| \leq C + |a|$ a.e.. Therefore

$$
e_n = \mathbb{E}[(a - Y)^2 \mid X = x_n] \leq (C + |a|)^2 < \infty.
$$

Consequently, $\sum_{n=1}^{\infty} a_n^2 e_n \leq (C + |a|)^2 \sum_{n=1}^{\infty} a_n^2 < \infty$. By summing (3.1.4), from $n = 1$ to $n = N$, we have that

$$
b_{n+1} = b_1 + \sum_{n=1}^{N} a_n^2 e_n - 2 \sum_{n=1}^{N} a_n d_n,
$$

which implies that $\sum_{n=1}^{\infty} a_n d_n$ converges, as

$$
\sum_{n=1}^{N} a_n d_n = \frac{1}{2}\left(b_1 - b_{N+1} + \sum_{n=1}^{N} a_n^2 e_n\right) \leq \frac{1}{2}\left(b_1 + \sum_{n=1}^{N} a_n^2 e_n\right)
$$

and $\sum_{n=1}^{\infty} a_n^2 e_n < \infty$. As a result, the sequence $(b_n)_n$ is convergent and its limit b is nonnegative:

$$
b = b_1 + \sum_{n=1}^{\infty} a_n^2 e_n - 2 \sum_{n=1}^{N} a_n d_n \geq b_1 + \sum_{n=1}^{\infty} a_n^2 e_n - \left(b_1 - \sum_{n=1}^{\infty} a_n^2 e_n\right) \geq 0.
$$

Let $(k_n)_n$ as in the Lemma's statement. Since $\sum_{n=1}^{\infty} a_n d_n < \infty$ and $d_n \geq k_n b_n$, we have that $\sum_{n=1}^{\infty} a_n k_n b_n \leq \sum_{n=1}^{\infty} a_n d_n < \infty$, from which we deduce that $b = 0$.[4] We showed that $(x_n)_n$ converges to θ in L_2. Convergence in probability follows by the Chebyshev inequality.[5] ■

The following lemma provides a sufficient condition, so that there exists some $(k_n)_n$ with the aforementioned properties:

[4] The sequence $(b_n)_n$ has a subsequence that converges to zero. Otherwise, there would exist some constant $M > 0$ such that $b_n \geq M$ for every n, which would force the sum $\sum_{n=1}^{\infty} a_n k_n b_n \geq M \sum_{n=1}^{\infty} a_n k_n = +\infty$ to diverge. As $(b_n)_n$ converges to b and has a null convergent subsequence, it has to be that $b = 0$.

[5] For every $\varepsilon > 0$, we have that $P[|x_n - \theta| \geq \varepsilon] \leq \frac{\|x_n - \theta\|_2^2}{\varepsilon^2} = \frac{b_n}{\varepsilon^2} \to 0$ as $n \to \infty$.

Lemma 3.1.3: *Under the notation of Lemma 3.1.2, let $(A_n)_n$ and $(k_n)_n$ be defined as*

$$A_n = |x_1 - \theta| + (C + |a|) \sum_{i=1}^{n-1} a_i, \tag{3.1.5}$$

$$k_n = \inf\left\{ \frac{M(x) - a}{x - \theta} : 0 < |x - \theta| < A_n \right\}. \tag{3.1.6}$$

Suppose that $\sum_{n=1}^{\infty} a_n = \infty$ and that there exists some $K > 0$ and $n_0 \in \mathbb{N}$ such that $k_n \geq \frac{K}{A_n}$ for every $n \geq n_0$. Then $(x_n)_n$ converges in probability to θ.

Proof. The functions $M(x) - a$ and $x - \theta$ have he same sign, so $(k_n)_n$ is a sequence of nonnegative numbers. Let P_n be the distribution of x_n. Then

$$\begin{aligned} d_n &= \mathbb{E}[(x_n - \theta)(M(x_n) - a)] = \int_{\mathbb{R}} (x - \theta)(M(x) - a) dP_n(x) \\ &= \int_{|x-\theta| \leq A_n} (x - \theta)(M(x) - a) dP_n(x) \\ &\geq \int_{|x-\theta| \leq A_n} k_n (x - \theta)^2 dP_n = k_n b_n, \end{aligned}$$

as $x_n - \theta$ is supported at $[-A_n, A_n]$.[6] So $(k_n)_n$ satisfies $d_n \geq k_n b_n$ for every n. It is left to show that $\sum_{n=1}^{\infty} a_n k_n = \infty$.

Let $n_0 \in \mathbb{N}$ such that $k_n \geq \frac{K}{A_n}$ for every $n \geq n_0$ and set $S_n = \sum_{i=1}^{n} a_i$. Let also $n_1 \in \mathbb{N}$ such that $(C + |a|) \sum_{i=1}^{n} a_i \geq |x_1 - \theta|$.[7] Then for every $n \geq n_3 = \max\{n_0, n_1\}$,

$$\begin{aligned} a_n k_n \geq \frac{a_n K}{A_n} &= \frac{a_n K}{|x_1 - \theta| + (C + |a|) \sum_{i=1}^{n} a_i} \\ &\geq \frac{a_n K}{2(C + |a|) \sum_{i=1}^{n} a_i} \\ &= \frac{K}{2(C + |a|)} \frac{a_n}{S_{n-1}}. \end{aligned}$$

By Lemma 3.1.1, we obtain that $\sum_{n=n_3}^{\infty} a_n k_n \geq \frac{K}{2(C+|a|)} \sum_{n=n_3}^{\infty} \frac{a_n}{S_{n-1}} = \infty$, and by Lemma 3.1.3 we conclude that $(x_n)_n$ converges to θ in probability. ∎

Definition 3.1.4: A sequence $(a_n)_n$ is said to be of type $1/n$ if it belongs to $\ell_2 \setminus \ell_1$, namely $(a_n)_n$ is square summable but not absolutely summable.

Theorem 3.1.5: *Suppose that there exists some $C > 0$ such that for every $x \in \mathbb{R}$, $P[|Y(x)| \leq C] = 1$. Additionally, $M(x) = \mathbb{E}[Y \mid X = x]$ is a nondecreasing function with $M(\theta) = a$ and $M'(\theta) > 0$. If $(a_n)_n$ is a nonnegative sequence of type $1/n$, then the sequence $(x_n)_n$ constructed by Robbins-Monro algorithm,*

$$x_{n+1} = x_n + a_n(a - y_n), \tag{3.1.7}$$

[6] One can easily confirm that $x_n - \theta = x_1 - \theta + \sum_{i=1}^{n-1} a_i(a - y_i)$. As we have already shown in the proof of the previous Lemma, $|a - Y| \leq C + |a|$ a.e., so $|x_n - \theta| \leq |x_1 - \theta| + (C + |a|) \sum_{i=1}^{\infty} a_i = A_n$ with probability one.

[7] Such a n_1 always exists, as the series $\sum_{n=1}^{\infty} a_n$ diverges.

converges in probability to θ, that is, $x_n \xrightarrow{P} \theta$.

Proof. By the differentiability of $M(x)$ at $x = \theta$, there exists some function $e(x)$ with the property that $M(x) = M(\theta) + M'(\theta)(x-\theta) + e(x)(x-\theta)$ and $\lim_{x\to\theta} e(x) = 0$. We pick a $\delta > 0$ such that $e(x) \geq -\frac{1}{2}M'(\theta) > 0$ for every $x \in (\theta-\delta, \theta+\delta)$.

For $\theta + \delta \leq x \leq \theta + A_n$,

$$\frac{M(x)-a}{x-\theta} \geq \frac{M(\theta+\delta)-a}{A_n} \geq \frac{\delta M'(\theta)}{2A_n}.$$

Similarly, the same holds for $\theta - A_n \leq x \leq \theta - \delta$, so

$$\frac{M(x)-a}{x-\theta} \geq \frac{\delta M'(\theta)}{2A_n} \tag{3.1.8}$$

for every x such that $|x-\theta| \leq \max\{A_n, \delta\}$. By picking δ sufficiently small, we can assure that $\max\{A_n, \delta\} = A_n$.[8]

We proved that $k_n = \inf\left\{\frac{M(x)-a}{x-\theta} : 0 < |x-\theta| < A_n\right\} \geq \frac{K}{A_n}$, where we set $K = \frac{\delta_1 M'(0)}{2} > 0$, so the convergence property follows from Lemma 3.1.3. ∎

The Robbins-Monro Algorithm	
Step 1	Pick a sequence $(a_n)_n \in \ell_2 \setminus \ell_1$ and fix some $x_1 \in \mathbb{R}$.
Step 2	Suppose that a sample $x_1, \ldots, x_{n-1}$ has been drawn. Draw a value y_n from the distribution $H(y \mid x_n)$ and set $x_{n+1} = x_n + a_n(a - y_n)$.
Step 3	Return to Step 2 and iterate.

Table 2: The Robbins-Monro Algorithm [RM51].

3.2 *Stochastic approximation algorithms*

Stochastic approximation algorithms are used in order to determine the fixed point of a (not necessarily linear) operator $H : X \to X$ defined on some vector space X. To obtain meaningful results, we will assume throughout that the given operator H has indeed a unique fixed point; for example, H could be a contraction defined on a Banach space.

The idea of using iterative algorithms for determining the fixed point of a function is not only old, but also quite fundamental. As a didactic example, recall one of the cornerstones of fixed point theorems, and real analysis in general, proven by Stefan Banach in 1922 [Ban22]:

[8] We have that $A_n \to \infty$, so there exists some $n_0 \in \mathbb{N}$ with $A_n \geq \delta$ for every $n \geq n_0$. We set $\delta_1 = \min\{\delta, A_1, \ldots, A_{n_0-1}\}$. Then for every $x \in (\theta-\delta_1, \theta+\delta_1)$, relation (3.1.8) holds and, additionally, $\max\{\delta, A_n\} = A_n$ for every $n \in \mathbb{N}$.

Definition 3.2.1: A function $H : (X, \rho) \to (X, \rho)$ defined on some metric space (X, ρ), is called a *β-contraction*, where β is a non-negative constant, if

$$\rho(Hx, Hy) \leq \beta\rho(x, y)$$

for every $x, y \in X$.

Theorem 3.2.2 (Banach's Fixed Point Theorem): *Let (X, ρ) be a complete metric space and $H : X \to X$ be a β-contraction for some $\beta \in (0, 1)$. Then H has a unique fixed point, namely there exists a unique $x_0 \in X$ with $Hx_0 = x_0$.*

The theorem is not just existential, but its proof also provides a way to approximate the fixed point. The idea behind it is simple, yet powerful. Pick any $x_0 \in X$ and follow the "orbit" of it through the iterations of H, namely the set $\text{orb}(x, H) = \{H^n x_0 : n \in \mathbb{N}\}$. Then, the sequence $(H^n x_0)_n$ converges to the unique fixed point of H. In particular, during the theorem's proof, one can show that for all n,

$$d\left(H^n x_0, x^*\right) \leq \frac{\beta^n}{1-\beta}\,(x_0, Hx_0) \longrightarrow 0, \tag{3.2.1}$$

thus also establishing the convergence rate to the fixed point x^*.

In Banach's theorem, we are essentially able to approach x^* by picking an arbitrary x_0, and then using the iteration

$$x_{n+1} = Hx_n \quad \text{for } n \in \mathbb{N}. \tag{3.2.2}$$

The good properties of the operator H and the underlying metric space, are strong enough to guarantee the convergence of the iterative algorithm for any initial point x_0.

Relying on the orbits of H will be crucial in our approach as well, however, in our setting we will not be able to invoke Banach's theorem directly, as some of its assumptions have to be relaxed. Most importantly, due to the presence of chance, the exact value of Hx_n may not be known, but instead only a value $Hx_n + w_n$, where w_n is an error term, can be observed. A modified iteration, similar to (3.2.2), would be

$$x_{n+1} = (1-\gamma)x_n + \gamma(Hx_n + w_n) \tag{3.2.3}$$

for some $\gamma \in (0, 1]$. For $\gamma = 1$ we obtain the direct analog of (3.2.2), whereas for $\gamma \in (0, 1)$, the latest iteration x_{n+1} is a convex combination of the old one x_n and the proposed value $Hx_n + w_n$.[9] If we can guarantee that the sequence $(x_n)_n$ generated by (3.2.3) is convergent to some $\tilde{x}$, if H is continuous and if $w_n \to 0$, then by taking limits in (3.2.3) we have that $\tilde{x} = H\tilde{x}$ and the algorithm converges to the desired fixed point.

[9] For $\gamma = 0$ the sequence $(x_n)_n$ is convergent, but it is also constant and equal to x_0, so it is not worth considering.

In this section we examine conditions which can assert this convergence. Our main goal is to establish the convergence of the Q-Learning algorithm. This proof, is far from straightforward, and will be presented in the next section. Essentially, all the results mentioned in this paragraph will be aimed towards this goal, but we should point out that they also have many other applications besides Q-Learning, as well as an intrinsic value on their own. Our treatment is based on Chapter 4 of the "Neuro-Dynamic Programming" textbook by Dimitri Bertsekas and John Tsitsiklis [BT96], whereas many of the results belong to John Tsitsiklis himself [Tsi94].

The operators under study, may not necessarily be contractions, but they satisfy a contraction-like property under some suitable norm. In our current discussion, the underlying space will be $\mathbb{R}^N$ and the suitable norm will be a weighted supremum norm:

Definition 3.2.3: Let $\xi \in \mathbb{R}^N$ with $\xi_i > 0$ for every $i = 1, \dots, N$. The function $\|\cdot\|_\xi : \mathbb{R}^N \to \mathbb{R}$ defined as

$$\|x\|_\xi = \max_{i=1,\dots,N} \frac{|x_i|}{\xi_i} \tag{3.2.4}$$

for $x \in \mathbb{R}^N$, is called the *weighted supremum norm induced by* ξ.

It is easy to see that $\|\cdot\|_\xi$ is indeed a norm[10] and by taking $\xi = e = (1, \dots, 1)$, the induced $\|\cdot\|_\xi$ is just the usual supremum norm. A known fact from functional analysis is that all norms in $\mathbb{R}^N$ are equivalent. This can be reformulated as follows: Any two normed spaces of the same finite dimension must be isomorphic [Arg04, Corollary 3.27].

It is not true, however, that they must also be isometrically isomorphic. For example, $(\mathbb{R}^n, \|\cdot\|_2)$ is never isometrically isomorphic to $(\mathbb{R}^n, \|\cdot\|_\infty)$ for $n \geq 2$, since the former is a Hilbert space, and the latter is not. An interesting property of weighted supremum spaces of the same dimension, is that they are all *isometrically isomorphic*. This is a useful remark that we will refer to later on:

Remark 3.2.4: Let $\xi \in \mathbb{R}^N$ be a strictly positive vector. Then $\left(\mathbb{R}^N, \|\cdot\|_\infty\right)$ is isometrically isomorphic to $\left(\mathbb{R}^N, \|\cdot\|_\xi\right)$. Indeed, let $S : \left(\mathbb{R}^N, \|\cdot\|_\xi\right) \to \left(\mathbb{R}^N, \|\cdot\|_\infty\right)$ defined as $S(x) = (x_i/\xi_i)_{i=1}^N$ for $x \in \mathbb{R}^N$. The operator S is linear and onto. Additionally,

$$\|Sx\|_\infty = \left\|(x_i/\xi_i)_{i=1}^N\right\|_\infty = \max_{i=1,\dots,N} \frac{|x_i|}{\xi_i} = \|x\|_\xi$$

for every $x \in \mathbb{R}^N$, so S is an isometry.

In a similar manner, $L = S^{-1} : \left(\mathbb{R}^N, \|\cdot\|_\infty\right) \to \left(\mathbb{R}^N, \|\cdot\|_\xi\right)$ defined as $L(x) = (\xi_i x_i)_{i=1}^N$ for $x \in \mathbb{R}^N$, is an isometric isomorphism. Both operators S and L will be used during the proof of the convergence of some stochastic approximation

[10] More generally, let $\xi : [0,1] \to \mathbb{R}$ be a continuous function with $\xi(x) > 0$ for every $x \in [0,1]$. Then $\|\cdot\|_\xi$ defined as $\|f\|_\xi = \sup_{x\in[0,1]} \frac{|f(x)|}{\xi(x)}$ induces a norm on $C[0,1]$ which is equivalent to its usual supremum norm.

algorithms, in our effort to reformulate the problem to an equivalent one, stated on an appropriate weighted supremum space (see Figure 1 for example).[11]

A class of functions that plays a central role in stochastic approximation, is the pseudo-contractions. These functions generalize the notion of a contraction and, naturally, they share some similar properties.

Definition 3.2.5: Let $(X, \|\cdot\|)$ be a normed space. A function $H : X \to X$ is called a *pseudo-contraction*, if there exists some $x^* \in X$ and some $\beta \in [0, 1)$ such that

$$\|Hx - x^*\| \leq \beta \|x - x^*\| \tag{3.2.5}$$

for every $x \in X$.

We collect a few remarks and examples which will help us demystify this notion:

Remark 3.2.6: Every pseudo-contraction H with x^* and β as in (3.2.5) has a unique fixed point at x^* and is continuous at x^*. Indeed, by setting $x = x^*$ in (3.2.5), we obtain that $\|Hx^* - x^*\| = 0$, thus $Hx^* = x^*$. In order to show that x^* is the only fixed point of H, suppose that x_0 is such that $Hx_0 = x_0$. Then by setting $x = x_0$ in (3.2.5), we obtain that $\|x_0 - x^*\| \leq \beta \|x_0 - x^*\|$, and since $\beta < 1$, this yields that $x_0 = x^*$.

For the continuity, suppose that $(x_n)_n$ is a sequence that converges to x^*. Then

$$\|Hx_n - x^*\| \leq b\|x_n - x^*\| \to 0,$$

so $Hx_n \to x^* = Hx^*$, as we wanted.

We mentioned that pseudo-contractions generalize the notion of a contraction, so one would expect that any contraction should be a pseudo-contraction as well. However, as we just showed, every pseudo-contraction has a fixed point, whereas, a contraction may not. So, in general, a contraction H need not be a pseudo-contraction. A necessary and sufficient condition for this implication to hold, is that H should possess a fixed point:

Remark 3.2.7: Suppose that H is a contraction with a fixed point x^*. Then H is a pseudo-contraction: Since H is a contraction, we have that $\|Hx - Hy\| \leq \beta\|x - y\|$ for every $x, y \in X$. Let $y = x^*$ be the fixed point of H. Then $\|Hx - x^*\| \leq \beta\|x - x^*\|$ for all x, thus H is a pseudo-contraction. In particular, every contraction is also a pseudo-contraction when X is a Banach space.[12]

It will often be useful, given a pseudo-contraction H with a fixed point at x^*, to "translate" it in a way that the resulting function H' is a pseudo-contraction with a fixed point at zero.

[11] On the other hand, none of the ℓ_p norms for $1 < p < \infty$ are isometrically isomorphic to any weighted supremum norm. This can easily be seen by examining the shape of the unit sphere under the various norms. In addition, and only for $N = 2$, the space $(\mathbb{R}^2, \|\cdot\|_1)$ is isometrically isomorphic to the usual supremum norm, thus to any weighted supremum norm as well.

[12] This is a consequence of Banach's fixed point theorem (see the discussion in the beginning of this section and Theorem 3.2.2).

Remark 3.2.8: Suppose that $H : X \to X$ is a pseudo-contraction with a fixed point at x^* and a constant $\beta \in [0,1)$. Then, the function $H' : X \to X$ defined as $H'(x) = H(x + x^*) - H(x^*)$ for $x \in X$, is a pseudo-contraction with a fixed point at zero, and with the same constant β: Let $x \in X$. Then $H'(0) = 0$ and

$$\begin{aligned}\|H'(x) - 0\| &= \|H(x + x^*) - H(x^*)\| \\ &= \|H(x + x^*) - x^* + x^* - H(x^*)\| \\ &= \|H(x + x^*) - x^*\| \leq \beta\|x\|.\end{aligned}$$

Example 3.2.9: An example of a contraction on a (non-complete) metric space that is not a pseudo-contraction: Let $H : (0,1) \to (0,1)$ defined as $H(x) = \frac{1}{2}x$. Clearly H is a contraction with $\beta = \frac{1}{2}$, but it is not a pseudo-contraction since H has no fixed points.

Example 3.2.10: An example of a pseudo-contraction which is not a contraction: Let $T : \mathbb{R} \to \mathbb{R}$ defined as follows:

$$T(x) = \begin{cases} \frac{1}{2}x, & x \neq 2012, 2016, \\ 4, & x = 2012, \\ 8, & x = 2016. \end{cases}$$

Then

$$\begin{aligned} |T(x) - T(0)| &\leq \frac{1}{2}|x - 0|, \\ |T(2012) - T(0)| = 4 &< \frac{2012}{2} \quad \text{and} \\ |T(2016) - T(0)| = 8 &< \frac{2016}{2}, \end{aligned}$$

which prove that T is a pseudo-contraction around zero. However, it is not a contraction as $|T(2012) - T(2016)| = 4$, which is never less or equal than 4β for any $\beta \in [0,1)$.

We can now study the convergence of some iterative algorithms, starting with the iteration

$$r_{n+1} = r_n + \gamma_n s_n, \tag{3.2.6}$$

where γ_n and s_n are random variables denoting the *step sizes* and the *update directions* respectively. We also let

$$\mathcal{F}_n = \sigma\left(r_0, \ldots, r_n, \gamma_0, \ldots, \gamma_n, s_0, \ldots, s_n\right), \tag{3.2.7}$$

denote the sigma-algebra generated by history of the algorithm up to the n-th step just before s_n is revealed. Although the following proposition, strictly speaking, does not assure convergence, it is a relevant convergence result on which some future arguments will be based:

Proposition 3.2.11: *Consider the iteration*

$$r_{n+1} = r_n + \gamma_n s_n, \tag{3.2.8}$$

where the step sizes $(\gamma_n)_n$ are such that $\sum_{n=0}^{\infty} \gamma_n(i) = \infty$ and $\sum_{n=0}^{\infty} \gamma_n(i)^2 < \infty$ for every $i = 1, \dots, N$. Let also $f : \mathbb{R}^N \to \mathbb{R}$ be a function for which

(a) $f(x) \geq 0$ for every $x \in \mathbb{R}^N$.

(b) The function ∇f is Lipschitz, that is, there exists some $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \text{for every } x, y \in \mathbb{R}^N. \tag{3.2.9}$$

(c) There exists some $C > 0$ such that

$$C\|\nabla f(r_n)\|^2 \leq -\langle \nabla f(r_n),\ \mathbb{E}[s_n \,|\, \mathcal{F}_n]\rangle \quad \text{for all } n \in \mathbb{N}. \tag{3.2.10}$$

(d) There exist $K_1, K_2 > 0$ such that

$$\mathbb{E}\left[\|s_n\|^2 \,|\, \mathcal{F}_n\right] \leq K_1 + K_2\|\nabla f(r_n)\|^2 \quad \text{for all } n \in \mathbb{N}. \tag{3.2.11}$$

Then, the sequence $(f(r_n))_n$ is convergent, the sequence $(\nabla f(r_n))_n$ converges to zero and every limit point of $(r_n)_n$ is a stationary point of f.[13]

Proof. We first show that for every $x, y \in \mathbb{R}^N$,

$$f(y) \leq f(x) + \langle \nabla f(x),\ y - x\rangle + \frac{L}{2}\|y - x\|^2. \tag{3.2.12}$$

Let $g(\lambda) = f\left(\lambda y + (1-\lambda)x\right)$. Then

$$\begin{aligned}
f(y) - f(x) = g(1) - g(0) &= \int_0^1 \frac{dg(\lambda)}{d\lambda} d\lambda = \int_0^1 \frac{df(\lambda y + (1-\lambda)x)}{d\lambda} d\lambda \\
&= \int_0^1 \left\langle y - x,\ \nabla f(\lambda y + (1-\lambda)x)\right\rangle d\lambda \\
&= \int_0^1 \left\langle y - x,\ \nabla f(\lambda y + (1-\lambda)x) - \nabla f(x) + \nabla f(x)\right\rangle d\lambda \\
&\leq \int_0^1 \langle y - x,\ \nabla f(x)\rangle d\lambda + \\
&\quad + \int_0^1 \|y - x\| \cdot \left\|\nabla f(\lambda y + (1-\lambda)x) - \nabla f(x)\right\| d\lambda \\
&\leq \int_0^1 \langle y - x,\ \nabla f(x)\rangle d\lambda + \int_0^1 L\lambda\|y - x\|^2 d\lambda
\end{aligned}$$

[13] The last statement means that for every $r \in \mathbb{R}^N$ for which there exists some subsequence of $(r_n)_n$ that converges to it (that is, r is a limit point of the sequence $(r_n)_n$), we have that $\nabla f(r) = 0$. Notice, however, that the Proposition does not guarantee the existence of any limit points of $(r_n)_n$ to begin with.

$$= \langle \nabla f(x),\ y - x \rangle + \frac{L}{2} \|y - x\|^2,$$

and (3.2.12) is established. By setting $x = x_n$ and $y = x_n + \gamma_n s_n$ in it, we obtain

$$f(r_{n+1}) \le f(r_n) + \gamma_n \langle \nabla f(r_n),\ s_n \rangle + \frac{L\gamma_n^2}{2} \|s_n\|^2,$$

and by taking expectations,

$$\begin{aligned}
\mathbb{E}[f(r_{n+1}) \,|\, \mathcal{F}_n] &\le \mathbb{E}[f(r_n) \,|\, \mathcal{F}_n] + \gamma_n \mathbb{E}[\langle \nabla f(r_n),\ s_n \rangle \,|\, \mathcal{F}_n] + \frac{L}{2} \mathbb{E}\left[\gamma_n^2 \|s_n\|^2 \,|\, \mathcal{F}_n\right] \\
&= f(r_n) + \gamma_n \langle \nabla f(r_n),\ \mathbb{E}[s_n \,|\, \mathcal{F}_n] \rangle + \frac{L\gamma_n^2}{2} \mathbb{E}[\|s_n\|^2 \,|\, \mathcal{F}_n] \\
&\le f(r_n) - \gamma_n C \|\nabla f(r_n)\|^2 + \frac{L\gamma_n^2}{2} \left(K_1 + K_2 \|\nabla f(r_n)\|^2\right) \\
&\le f(r_n) - \gamma_n \|\nabla f(r_n)\|^2 \left(C - \frac{L\gamma_n}{2} K_2\right) + \frac{K_1 L \gamma_n^2}{2}, \qquad (3.2.13)
\end{aligned}$$

due to Proposition B.3.3. We set

$$X_n = \begin{cases} \gamma_n \left(C - \frac{LK_2\gamma_n}{2}\right) \|\nabla f(r_n)\|^2, & \text{if } LK_2\gamma_n \le 2C, \\ 0, & \text{otherwise,} \end{cases} \qquad (3.2.14)$$

$$Z_n = \begin{cases} \frac{LK_1\gamma_n^2}{2}, & \text{if } LK_2\gamma_n \le 2C, \\ \frac{LK_1\gamma_n^2}{2} - \gamma_n \left(C - \frac{LK_2\gamma_n}{2}\right) \|\nabla f(r_n)\|^2, & \text{otherwise,} \end{cases} \qquad (3.2.15)$$

and (3.2.13) can written as

$$\mathbb{E}[f(r_{n+1}) \,|\, \mathcal{F}_n] = f(r_n) - X_n + Z_n. \qquad (3.2.16)$$

Both $(X_n)_n$ and $(Z_n)_n$ consist of non-negative random variables with the property that X_n and Z_n are $\mathcal{F}_n$-measurable for every $n \in \mathbb{N}$. Since $\gamma_n \to 0$, we can pick an $n_0 \in \mathbb{N}$ such that $Z_n = \frac{LK_1\gamma_n^2}{2}$ for every $n \ge n_0$, thus $\sum_{n=1}^{\infty} Z_n < \infty$. By the Supermartingale Convergence Theorem (see p. 135), we obtain that $(f(r_n))_n$ converges almost surely to a non-negative random variable, and that $\sum_{n=1}^{\infty} X_n < \infty$ as well. So the first claim has been proven.

We pick an $n_1 \in \mathbb{N}$ such that $\frac{LK_2\gamma_n}{2} \le \frac{C}{2}$ for every $n \ge n_1$. Then

$$X_n = \gamma_n \left(C - \frac{LK_2\gamma_n}{2}\right) \|\nabla f(r_n)\|^2 \ge \frac{C}{2} \gamma_n \|\nabla f(r_n)\|^2$$

for every $n \ge n_1$ and $+\infty > \sum_{n=1}^{\infty} X_n \ge \frac{C}{2} \sum_{n=1}^{\infty} \gamma_n \|\nabla f(r_n)\|^2$, so

$$\sum_{n=1}^{\infty} \gamma_n \|\nabla f(r_n)\|^2 < \infty. \qquad (3.2.17)$$

This implies that $\liminf_n \|\nabla f(r_n)\|^2 = 0$. [14] We will show that $\|\nabla f(r_n)\| \to 0$.

Let $\varepsilon > 0$. We say that the interval of natural numbers $[n, n^*]$ is an *upcrossing interval from $\varepsilon/2$ to ε*, if $\|\nabla f(r_n)\| < \frac{\varepsilon}{2}$, $\|\nabla f(r_{n^*})\| > \varepsilon$ and $\frac{\varepsilon}{2} \le \|\nabla f(r_k)\| \le \varepsilon$, for every $k \in (n, n^*)$. We set $w_n = s_n - \mathbb{E}[s_n|\mathcal{F}_n]$. By the projection property of the conditional expectation (see Proposition B.3.2), each w_n is orthogonal to $\mathbb{E}[s_n \,|\, \mathcal{F}_n]$, so by the Pythagorean Theorem,

$$\|w_n\|^2 + \|\mathbb{E}[s_n \,|\, \mathcal{F}_n]\|^2 = \|w_n + \mathbb{E}[s_n \,|\, \mathcal{F}_n]\|^2 = \|s_n\|^2,$$

and after conditioning on $\mathcal{F}_n$ we obtain that

$$\begin{aligned} &\mathbb{E}[\|w_n\|^2 \,|\, \mathcal{F}_n] + \mathbb{E}\left[\|\mathbb{E}[s_n \,|\, \mathcal{F}_n]\|^2 \,|\, \mathcal{F}_n\right] = \mathbb{E}[\|s_n\|^2 \,|\, \mathcal{F}_n] \quad \Rightarrow \\ &\qquad \mathbb{E}[\|w_n\|^2 \,|\, \mathcal{F}_n] + \|\mathbb{E}[s_n \,|\, \mathcal{F}_n]\|^2 \le K_1 + K_2\|\nabla f(r_n)\|^2 \end{aligned} \tag{3.2.18}$$

for every n. Let $(\mathcal{X}_n)_n$ be defined as

$$\mathcal{X}_n = \begin{cases} 1, & \text{if } \|\nabla f(r_n)\| < \varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

Each $\mathcal{X}_n$ is clearly $\mathcal{F}_n$-measurable. We also define the sequence $(U_n)_n$ as follows: $U_n = \sum_{k=0}^{n-1} \mathcal{X}_k \gamma_k w_k$. We will show that $(U_n)_n$ converges almost surely whenever $\sum_{n=0}^{\infty} \gamma_n^2 \le A$ almost surely for some deterministic constant A:

Each U_n is $\mathcal{F}_n$-measurable with

$$\mathbb{E}[\mathcal{X}_n \gamma_n w_n \,|\, \mathcal{F}_n] = \mathcal{X}_n \gamma_n \mathbb{E}[w_n \,|\, \mathcal{F}_n] = 0,$$

therefore

$$\mathbb{E}[U_{n+1} \,|\, \mathcal{F}_n] = \mathbb{E}[U_n + \mathcal{X}_n \gamma_n w_n \,|\, \mathcal{F}_n] = \mathbb{E}[U_n \,|\, \mathcal{F}_n] = U_n$$

for all n. When $\mathcal{X}_n = 0$,

$$\mathbb{E}\left[\|U_{n+1}\|^2 \,|\, \mathcal{F}_n\right] = \mathbb{E}\left[\|U_n\|^2 \,|\, \mathcal{F}_n\right] = \|U_n\|^2, \tag{3.2.19}$$

whereas when $\mathcal{X}_n = 1$,

$$\begin{aligned} \mathbb{E}\left[\|U_{n+1}\|^2 \,|\, \mathcal{F}_n\right] &= \mathbb{E}\left[\|U_n + \gamma_n\|w_n\|^2 \,|\, \mathcal{F}_n\right] \\ &= \mathbb{E}\left[\|U_n\|^2 + \gamma_n^2\|w_n\|^2 + 2\gamma_n\langle U_n,\ w_n\rangle \,|\, \mathcal{F}_n\right] \\ &= \|U_n\|^2 + \gamma_n^2 \mathbb{E}\left[\|w_n\|^2 \,|\, \mathcal{F}_n\right] + 2\gamma_n\langle U_n,\ \mathbb{E}[w_n \,|\, \mathcal{F}_n]\rangle \\ &= \|U_n\|^2 + \gamma_n^2 \mathbb{E}\left[\|w_n\|^2 \,|\, \mathcal{F}_n\right] \\ &\le \|U_n\|^2 + \gamma_n^2\left(K_1 + K_2\|\nabla f(r_n)\|^2\right) \\ &\le \|U_n\|^2 + \gamma_n^2\left(K_1 + \varepsilon^2 K_2\right), \end{aligned} \tag{3.2.20}$$

[14] Otherwise there would exist some $\varepsilon > 0$ and $n_2 \in \mathbb{N}$ such that $\|\nabla f(r_n)\| \ge \varepsilon$ for every $n \ge n_2$, which would in turn force the series $\sum_{n=1}^{\infty} \gamma_n \|\nabla f(r_n)\|^2 \ge \varepsilon \sum_{n=n_2}^{\infty} \gamma_n = +\infty$ to diverge.

by (3.2.18). By taking expectations in (3.2.20),

$$\mathbb{E}\left[\mathbb{E}\left[\|U_{n+1}\|^2 \,|\, \mathcal{F}_n\right]\right] = \mathbb{E}\left[\|U_{n+1}\|^2\right] \leq \mathbb{E}\left[\|U_n\|^2\right] + \left(K_1 + \varepsilon^2 K_2\right)\mathbb{E}\left[\gamma_n^2\right],$$

so

$$\sum_{n=0}^{N-1} \mathbb{E}\left[\|U_{n+1}\|^2\right] \leq \sum_{n=0}^{N-1} \mathbb{E}\left[\|U_n\|^2\right] + \left(K_1 + \varepsilon^2 K_2\right)\mathbb{E}\left[\sum_{n=0}^{N-1} \gamma_n^2\right], \quad \text{so}$$

$$\mathbb{E}\left[\|U_N\|^2\right] \leq \left(K_1 + \varepsilon^2 K_2\right)\mathbb{E}\left[\sum_{n=0}^{N-1} \gamma_n^2\right] \leq A\left(K_1 + \varepsilon^2 K_2\right).$$

By taking suprema in the last expression,

$$\sup_n \mathbb{E}[\|U_n\|] \leq \sup_n \mathbb{E}\left[1 + \|U_n\|^2\right] \leq 1 + A\left(K_1 + \varepsilon^2 K_2\right),$$

and the almost sure convergence of $(U_n)_n$ is asserted by the Martingale Convergence Theorem (Theorem B.3.5).

We now extend the claim we just proved, for the case where $\sum_{n=0}^{\infty} \gamma_n^2$ is not bounded by the same determninistic constant, but it is merely convergent almost surely. We will show that $(U_n)_n$ converges almost surely when $\sum_{n=0}^{\infty} \gamma_n^2 < \infty$ almost surely.

For every $k \in \mathbb{N}$ we define the process $(U_n^k)_n$ as follows:

$$U_n^k = \begin{cases} U_n, & \text{when } \sum_{i=0}^{n-1} \gamma_i^2 \leq k, \\ U_{n_0}, & \text{when } n \geq n_0, \end{cases}$$

where $n_0 = \min\left\{n : \sum_{i=0}^{n-1} \gamma_i^2 > k\right\}$. The sequence $(U_n^k)_n$ is either eventually constant, if there exists such an n_0, or it is equal to $(U_n)_n$ if $\sum_{i=0}^{\infty} \gamma_i^2 \leq k$. In either case, it converges almost surely. Set

$$\Omega_k = \left\{\omega \in \Omega : \left(U_n^k(\omega)\right)_n \text{ does not converge}\right\} \quad \text{and} \quad \Omega' = \Omega \setminus \bigcup_{k=1}^{\infty} \Omega_k.$$

Then $P(\Omega') = 1$ and for every $\omega \in \Omega'$, $(U_n^k(\omega))_n$ converges for every k. Let also

$$\Omega'' = \left\{\omega \in \Omega' : \sum_{i=0}^{\infty} \gamma_i(\omega) < \infty\right\}.$$

Then $P(\Omega'') = 1$ and for every $\omega \in \Omega''$, there exists some $k_\omega \in \mathbb{N}$ with $\sum_{i=0}^{\infty} \gamma_i(\omega)^2 \leq k_\omega$, so $U_n^k(\omega) = U_n(\omega)$ for every $k \geq k_\omega$ and since $\omega \in \Omega'$, the sequence $(U_n^k(\omega))_n$ converges for every $k \geq k_\omega$. This implies that $(U_n(\omega))_n$ converges for every $\omega \in \Omega''$, as we wanted.

Using the previous claim, we will show that a sample path may only have finitely many upcrossing intervals from $\varepsilon/2$ to ε. Suppose not. Let $\left([t_k, t_k']\right)_k$ be a sequence

of upcrossing intervals. We have that $\mathcal{X}_n = 1$ for $n \in [t_k, t_k')$, so $\sum_{n=t_k}^{t_k'-1} \gamma_n w_n \leq \sum_{n=t_k}^{\infty} \mathcal{X}_n \gamma_n w_n$ with

$$\lim_{k\to\infty} \sum_{n=t_k}^{t_k'-1} \gamma_n w_n \leq \lim_{k\to\infty} \sum_{n=t_k}^{\infty} \mathcal{X}_n \gamma_n w_n = 0 \tag{3.2.21}$$

almost surely. By the triangle inequality,

$$\begin{aligned} \left\|\nabla f(r_{t_k+1})\right\| - \left\|\nabla f(r_{t_k})\right\| &\leq \left\|\nabla f(r_{t_k+1}) - \nabla f(r_{t_k})\right\| \leq L \left\|r_{t_k+1} - r_{t_k}\right\| \\ &= L \left\|\gamma_{t_k} s_{t_k}\right\| \leq L\gamma_{t_k} \left\|w_{t_k}\right\| + L\gamma_{t_k} \left\|\mathbb{E}\left[s_{t_k} \mid \mathcal{F}_{t_k}\right]\right\| \\ &\leq L\gamma_{t_k} \left\|w_{t_k}\right\| + L\gamma_{t_k} \left(K_1 + K_2\varepsilon^2\right) \end{aligned}$$

by (3.2.18), so $\|\nabla f(r_{t_k+1})\| - \|\nabla f(r_{t_k})\| \xrightarrow[k\to\infty]{} 0$, and since $\|\nabla f(r_{t_k+1})\| > \varepsilon$, there exists some k_0 such that $\|\nabla f(r_{t_k})\| \geq \varepsilon/4$ for every $k \geq k_0$. Additionally,

$$\begin{aligned} \frac{\varepsilon}{2} &\leq \left\|\nabla f\left(r_{t_k'}\right)\right\| - \left\|\nabla f\left(r_{t_k}\right)\right\| \leq \left\|\nabla f\left(r_{t_k'}\right) - f\left(r_{t_k}\right)\right\| \\ &\leq L \left\|r_{t_k'} - r_{t_k}\right\| = L \left\|\sum_{n=t_k}^{t_k'-1} \gamma_n s_n\right\| \\ &\leq L \sum_{n=t_k}^{t_k'-1} \gamma_n \left\|\mathbb{E}\left[s_n \mid \mathcal{F}_n\right]\right\| + L \left\|\sum_{n=t_k}^{t_k'-1} \gamma_n w_n\right\| \\ &\leq L \sum_{n=t_k}^{t_k'-1} \gamma_n \left(1 + K_1 + K_2\varepsilon^2\right) + L \left\|\sum_{n=t_k}^{t_k'-1} \gamma_n w_n\right\|, \end{aligned}$$

with the second term converging to zero as $k \to \infty$, by (3.2.21). Therefore,

$$\liminf_k \sum_{n=t_k}^{t+k'} \gamma_n \geq \frac{\varepsilon}{2L\left(1 + K_1 + K_2\varepsilon^2\right)}. \tag{3.2.22}$$

On the other hand, for $n \in [t_k, t_k')$, we have that $\|\nabla f(r_n)\| \geq \frac{\varepsilon}{4}$, so

$$\begin{aligned} \liminf_k \sum_{n=t_k}^{t+k'} \gamma_n \|\nabla f(r_n)\|^2 &\geq \frac{\varepsilon^3}{32L\left(1 + K_1 + K_2\varepsilon^2\right)} \text{ and} \\ \sum_{n=0}^{\infty} \gamma_n \|\nabla f(r_n)\|^2 &\geq \sum_{n \in \cup_k [t_k, t_k - 1)} \gamma_n \|\nabla f(r_n)\|^2 = +\infty, \end{aligned}$$

which contradicts (3.2.17). We showed that for every $\omega \in \Omega$, there exist finitely many upcrossing intervals from $\varepsilon/2$ to ε. This means that for every $\varepsilon > 0$, there exists some $n_\varepsilon^* \in \mathbb{N}$ with $0 \leq \|\nabla f(r_n)\| \leq \varepsilon$ for every $n \geq n_\varepsilon^*$, implying that $\nabla f(r_n) \to 0$.

Lastly, if r_0 is a limit point of $(r_n)_n$, there exists some subsequence of it with $r_{k_n} \to r_0$ and by the (Lipschitz) continuity of ∇f, we have that $\nabla f(r_{k_n}) \to 0 = \nabla f(r_0)$, so r_0 is a stationary point of f. ■

Remark 3.2.12: The conclusion of the previous proposition still holds if assumption (d) is replaced by the weaker one:

(d') There exist a $K_2 > 0$ and a sequence of random variables $(A_n)_n$ which is bounded a.s., such that

$$\mathbb{E}\left[\|s_n\|^2 \,|\, \mathcal{F}_n\right] \leq A_n + K_2\|\nabla f(r_n)\|^2 \quad \text{for all } n \in \mathbb{N}. \tag{3.2.23}$$

Indeed, let Ω' denote the set of all ω's for which $(A_n(\omega))_n$ is bounded. We proceed as in the proof of Proposition 3.2.11, except that $(Z_n)_n$ will be now defined as

$$Z_n = \begin{cases} \frac{LA_n\gamma_n^2}{2}, & \text{if } LK_2\gamma_n \leq 2C, \\ \frac{LA_n\gamma_n^2}{2} - \gamma_n\left(C - \frac{LK_2\gamma_n}{2}\right)\|\nabla f(r_n)\|^2, & \text{otherwise.} \end{cases} \tag{3.2.24}$$

By the convergence of $(\gamma_n)_n$ to zero, there exists some n_0 such that $Z_n = \frac{LA_n\gamma_n^2}{2}$ for every $n \geq n_0$. In particular, for every $\omega \in \Omega'$, by the boundedness of $(A_n(\omega))_n$, we have that $Z_n(\omega) = \frac{LA_n(\omega)\gamma_n^2}{2} \leq \frac{LM_\omega\gamma_n^2}{2}$ for some $M_\omega > 0$, thus $\sum_{n=1}^{\infty} Z_n(\omega) < \infty$ for all $\omega \in \Omega'$. The rest of the proof remains unchanged.

Corollary 3.2.13: *Consider the recursion*

$$r_{n+1} = (1-\gamma_n)r_n + \gamma_n w_n, \tag{3.2.25}$$

where

(a) the step sizes $(\gamma_n)_n$ are such that $\sum_{n=1}^{\infty}\gamma_n(i) = \infty$ and $\sum_{n=1}^{\infty}\gamma_n(i)^2 < \infty$ for every $i = 1, \ldots, N$.

(b) The noise terms $(w_n)_n$ are such that $\mathbb{E}[w_n \,|\, \mathcal{F}_n] = 0$ and $\mathbb{E}[w_n^2 \,|\, \mathcal{F}_n] \leq A_n$, where $(A_n)_n$ is a sequence of random variables which is bounded with probability one, with each A_n being $\mathcal{F}_n$-measurable.

Then $(r_n)_n$ converges to zero with probability one.

Proof. By setting $s_n = w_n - r_n$ into (3.2.13), we can can obtain (3.2.25) as a special case of it, $r_{n+1} = r_n + \gamma_n s_n$. Set $f(r) = r^2 \geq 0$, with $\nabla f(r) = 2r$. We will confirm that all of the assumptions of Proposition 3.2.11 are met. Clearly, ∇f is a 2-Lipschitz function and

$$-\left\langle \nabla f(r_n),\, \mathbb{E}[w_n - r_n \,|\, \mathcal{F}_n]\right\rangle = -\langle 2r_n,\, -r_n\rangle = 2\|r_n\|^2, \text{ so}$$
$$c\|\nabla f(r_n)\|^2 = 4c\|r_n\|^2 \leq 2\|r_n\|^2$$

is satisfied for $c = \frac{1}{2}$. Additionally, by Proposition B.3.3, we have that for every n, $\mathbb{E}[\langle w_n,\, r_n\rangle \,|\, \mathcal{F}_n] = \langle r_n,\, \mathbb{E}[w_n \,|\, \mathcal{F}_n]\rangle$, so

$$\mathbb{E}\left[\|s_n\|^2 \,|\, \mathcal{F}_n\right] = \mathbb{E}\left[\|w_n - r_n\|^2 \,|\, \mathcal{F}_n\right] = \mathbb{E}\left[\|w_n\|^2 + \|r_n\|^2 + 2\langle w_n,\, r_n\rangle \,|\, \mathcal{F}_n\right]$$

$$\begin{aligned}&\leq A_n + \|r_n\|^2 + 2\mathbb{E}\left[\langle w_n,\, r_n\rangle \,|\, \mathcal{F}_n\right]\\ &= A_n + \frac{1}{4}\|2r_n\|^2 + 2\langle r_n,\, \mathbb{E}\left[w_n \,|\, \mathcal{F}_n\right]\rangle = A_n + \frac{1}{4}\|\nabla f(r_n)\|^2\end{aligned}$$

with probability one. By Proposition 3.2.11 and Remark 3.2.12, we obtain that $\lim_{n\to\infty} \nabla f(r_n) = \lim_{n\to\infty} 2r_n = 0$ with probability one. ■

We now state one of the main tools that will be used in the stochastic approximation for pseudo-contractions. It does not assert the convergence of $(r_n)_n$, but merely its boundedness, so it is more of an auxiliary result.

Proposition 3.2.14: *We consider the iteration*

$$r_{n+1} = (1-\gamma_n)r_n + \gamma_n(H_n r_n + w_n + u_n), \tag{3.2.26}$$

where

(a) the step sizes $(\gamma_n)_n$ are such that $\sum_{n=1}^{\infty} \gamma_n(i) = \infty$ and $\sum_{n=1}^{\infty} \gamma_n(i)^2 < \infty$ for every $i = 1, \ldots, N$.

(b) The noise terms $(w_n)_n$ have the properties that

$$\mathbb{E}\left[w_n(i) \,|\, \mathcal{F}_n\right] = 0 \quad \text{and} \quad \mathbb{E}\left[w_n(i)^2 \,|\, \mathcal{F}_n\right] \leq A + B\|r_n\|^2.$$

(c) Each H_n is a map $H_n : \mathbb{R}^N \to \mathbb{R}^N$, and there exist $\xi \in \mathbb{R}^N_{++}$, $\beta \in [0,1)$ and $D > 0$ such that $\|H_n r_n\|_\xi \leq \|r_n\|_\xi + D$.

(d) There exists a sequence of non-negative random variables $(\theta_n)_n$ which converges to zero with probability one, such that

$$\|u_n\|_\infty \leq \theta_n\left(1 + \|r_n\|_\xi\right)$$

for every $n \in \mathbb{N}$.

Then, the sequence $(r_n)_n$ is bounded with probability one.

Proof. Let us consider first the case where $\xi = (1, \ldots, 1)$ and $\|\cdot\|_\xi$ is just the supremum norm. We pick a $G \geq 1$ such that $\beta G + D < G$ and an h such that $\beta G + D = hG$. Then $\beta < h < 1$ and we can pick an $\varepsilon > 0$ such that $(1+\varepsilon)h = 1$. We define a sequence $(G_n)_n$ as follows: Set $G_0 = \max\{\|r_0\|_\infty, G\}$ and suppose that G_n has been defined. Then

$$G_{n+1} = \begin{cases} G_n, & \text{if } \|r_{n+1}\|_\infty \leq (1+\varepsilon)G_n,\\ G_0(1+\varepsilon)^k, & \text{otherwise,}\end{cases} \tag{3.2.27}$$

where $k \in \mathbb{N}$ is such that $G_0(1+\varepsilon)^{k-1} < \|r_{n+1}\|_\infty \leq G_0(1+\varepsilon)^k = G_{n+1}$. We notice that $\|r_n\|_\infty \leq (1+\varepsilon)G_n$ for every $n \in \mathbb{N}$ and that $\|r_n\|_\infty \leq G_n$ whenever $G_{n-1} < G_n$. Additionally, $(G_n)_n$ is increasing: If $G_{n+1} \neq G_n$, then $G_{n+1} = G_0(1+\varepsilon)^k \geq \|r_{n+1}\|_\infty > (1+\varepsilon)G_n > G_n$.

Since $\beta\varepsilon + h < h\varepsilon + h < 1$, we can pick a $\theta^* \geq 0$ such that $\beta\varepsilon + h + \theta^*(2+\varepsilon) \leq 1$. We also pick an $n^* \in \mathbb{N}$ such that $\theta_n \leq \theta^*$ with probability one for every $n \geq n^*$. Then

$$\begin{aligned}
\|H_n r_n\|_\infty + \theta_n(\|r_n\|_\infty + 1) &\leq \beta\|r_n\|_\infty + D + \theta^*(1 + (1+\varepsilon)G_n) \\
&\leq \beta(1+\varepsilon)G_n + (h-\beta)G + \theta^*(1 + (1+\varepsilon)G_n) \\
&\leq \beta(1+\varepsilon)G_n + (h-\beta)G_n + \theta^*(1 + (1+\varepsilon)G_n) \\
&= (\beta\varepsilon + h)G_n + \theta^*(2+\varepsilon)G_n \\
&\leq G_n. \qquad (3.2.28)
\end{aligned}$$

Let $w'_n = \frac{w_n}{G_n}$. We have that

$$\begin{aligned}
\mathbb{E}[w'_n(i) \mid \mathcal{F}_n] &= \frac{\mathbb{E}[w_n(i) \mid \mathcal{F}_n]}{G_n} = 0 \text{ and} \\
\mathbb{E}[w'_n(i)^2 \mid \mathcal{F}_n] &= \frac{\mathbb{E}[w_n^2(i) \mid \mathcal{F}_n]}{G_n^2} \leq \frac{A + B\|r_n\|^2}{G_n^2} \leq \frac{A + B(1+\varepsilon)^2 G_n^2}{G_n^2} \\
&\leq A + B(1+\varepsilon)^2.
\end{aligned}$$

We define the following families of recursions: For every $n_0 \in \mathbb{N}$, let $W'_{n_0,n_0} = 0$ and

$$W'_{n+1,n_0} = (1-\gamma_n)W'_{n,n_0} + \gamma_n w'_n \text{ for } n \geq n_0.$$

We will show that for every $\delta > 0$, there exists some $n_0 \in \mathbb{N}$ such that

$$\left\| W'_{n,n_0} \right\|_\infty \leq \delta \qquad (3.2.29)$$

for every $n \geq n_0$ almost surely. By Corollary 3.2.13, we have that $W'_{n,0}(i) \to 0$ almost surely, for every $i = 1, \ldots, N$. Using induction we can show that

$$W'_{n,0} = \prod_{k=n_0}^{n-1}(1-\gamma_k)W'_{n_0,0} + W'_{n,n_0} \qquad (3.2.30)$$

for every $n \geq n_0$. For $n = n_0$ it holds trivially. Suppose that $W'_{n,0} = \prod_{k=n_0}^{n-1}(1-\gamma_k)W'_{n_0,0} + W'_{n,n_0}$ for some $n \geq n_0$. Then

$$\begin{aligned}
W'_{n+1,0} &= (1-\gamma_n)W'_{n,0} + \gamma_n w'_n \\
&= (1-\gamma_n)\prod_{k=n_0}^{n-1}(1-\gamma_k)W'_{n_0,0} + (1-\gamma_n)W_{n,n_0} + \gamma_n w'_n \\
&= \prod_{k=n_0}^{n}(1-\gamma_k)W'_{n_0,0} + W_{n+1,n_0},
\end{aligned}$$

and the induction is complete. By (3.2.30) we have that for n_0 large enough so that $\|\gamma_n\|_\infty \leq 1$, $\|W_{n_0,0}\|_\infty < \frac{\delta}{2}$ and $\|W_{n,0}\|_\infty < \frac{\delta}{2}$ all hold with probability one for every $n \geq n_0$,

$$\begin{aligned}\|W'_{n,n_0}\|_\infty &= \left\| W'_{n,0} - \prod_{k=n_0}^{n-1}(1-\gamma_k) W'_{n_0,0} \right\|_\infty \\ &\le \|W'_{n,0}\|_\infty + \prod_{k=n_0}^{n-1}(1-\gamma_k)\,\|W'_{n_0,0}\|_\infty \\ &\le \frac{\delta}{2} + \frac{\delta}{2} = \delta\end{aligned}$$

for every $n \ge n_0$.

Pick an $n_0 \in \mathbb{N}$ such that $\|r_{n_0}\|_\infty \le G_{n_0}$, $\|W'_{n,n_0}\|_\infty \le \varepsilon$, $\|\gamma_n\| \le 1$ and $\theta_n \le \theta^*$ all hold with probability one for every $n \ge n_0$. We will show inductively that

$$\begin{aligned} G_n = G_{n_0} \quad &\text{and} \\ -G_{n_0}(1+\varepsilon) \le -G_{n_0} + W'_{n,n_0} G_{n_0} \le r_n \le G_{n_0} + W'_{n,n_0} G_{n_0} &\le G_{n_0}(1+\varepsilon)\end{aligned} \tag{3.2.31}$$

for every $n \ge n_0$. For $n = n_0$, (3.2.31) becomes

$$-G_{n_0}(1+\varepsilon) \le -G_{n_0} \le r_n \le G_{n_0} + \le G_{n_0}(1+\varepsilon),$$

which clearly holds. Suppose that (3.2.31) both hold for some $n \ge n_0$. Then

$$\begin{aligned} r_{n+1} &= (1-\gamma_n) r_n + \gamma_n (H_n r_n + w_n + u_n) \\ &\le (1-\gamma_n)(G_{n_0} + W'_{n,n_0} + G_{n_0}) + \gamma_n w'_n G_n + \gamma_n (H_n r_n + \theta_n(\|r_n\|_\infty + 1)) \\ &\le G_{n_0} + G_{n_0}\left((1-\gamma_n) W'_{n,n_0} + w'_n \gamma_n\right) \\ &= G_{n_0} + G_{n_0} W'_{n+1,n_0}, \end{aligned}$$

where we used the induction hypothesis on r_n and the fact that $G_n = G_{n_0}$. In a similar manner we obtain the other half of (3.2.31). Additionally, $\|r_{n+1}\|_\infty \le G_{n_0}(1+\varepsilon)$ implies that $G_n = G_{n_0}$. We showed that $G_n = G_{n_0}$ for every $n \ge n_0$. The boundedness of $(r_n)_n$ then follows from the relation $\|r_n\|_\infty \le (1+\varepsilon) G_n = (1+\varepsilon) G_{n_0}$ which holds for every $n \ge n_0$.

We now examine the general case, where ξ is strictly positive but not necessarily equal to $e = (1, \dots, 1)$. Using the isometries S and L between $(\mathbb{R}^N, \|\cdot\|_\infty)$ and $(\mathbb{R}^N, \|\cdot\|_\xi)$, as introduced in Remark 3.2.4, we can reformulate the problem as follows (see also Figure 1):

We set $r'_n = S r_n$, $H'_n = S \circ H_n \circ L$, $w'_n = S w_n$ and $u'_n = S u_n$ for every $n \in \mathbb{N}$. Then, the elements r'_n belong to $(\mathbb{R}^N, \|\cdot\|_\infty)$ and by applying the operator S to the iteration (3.2.26) we obtain that:

$$\begin{aligned} r_{n+1} &= (1-\gamma_n) r_n + \gamma_n (H_n r_n + w_n + u_n) && \Rightarrow \\ S r_{n+1} &= (1-\gamma_n) S r_n + \gamma_n (S H_n r_n + S w_n + S u_n) && \Rightarrow \\ r'_{n+1} &= (1-\gamma_n) r'_n + \gamma_n (S H_n L (S r_n) + w'_n + u'_n) && \Rightarrow \\ r'_{n+1} &= (1-\gamma_n) r'_n + \gamma_n (H'_n(r'_n) + w'_n + u'_n), \end{aligned} \tag{3.2.32}$$

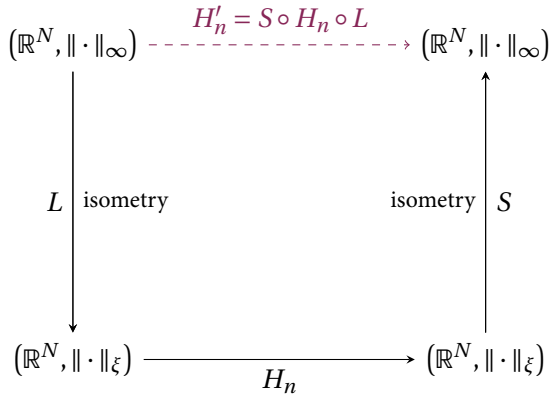

Figure 1: Reformulating the problem so that the induced operator H'_n on $(\mathbb{R}^n, \|\cdot\|_\infty)$ has the same properties as the original one H_n defined on $(\mathbb{R}^n, \|\cdot\|_\xi)$.

so the newly introduced elements r'_n, H'_n, w'_n and u'_n satisfy the same relation (3.2.26). Furthermore,

$$\begin{aligned}
\mathbb{E}\left[w'_n(i) \,|\, \mathcal{F}_n\right] &= \frac{1}{\xi_i}\mathbb{E}\left[w_n(i) \,|\, \mathcal{F}_n\right] = 0,\\
\mathbb{E}\left[w'_n(i)^2 \,|\, \mathcal{F}_n\right] &= \frac{1}{\xi_i^2}\mathbb{E}\left[w_n(i)^2 \,|\, \mathcal{F}_n\right] \le \max_{i=1,\dots,n}\left\{\frac{1}{\xi_i^2}\right\}\mathbb{E}\left[w_n(i)^2 \,|\, \mathcal{F}\right]\\
&\le C(A + B\|r_n\|_\xi) \le A' + B'\|r_n\|_\infty,\\
\|H'_n r'_n\|_\infty &= \|SH_nLr'_n\|_\infty = \|H_nLr'_n\|_\xi \le \beta\|Lr'_n\|_\xi + D\\
&= \beta\|r'_n\|_\infty + D,\\
\|u'_n\|_\infty &= \|u_n\|_\xi \le \theta_n(1 + \|r_n\|_\xi) \le \theta_n(1 + C\|r_n\|_\infty),
\end{aligned}$$

the last inequality holding by the equivalency of the $\|\cdot\|_\xi$ and $\|\cdot\|_\infty$ norms. If $C \le 1$, then $\|u'_n\|_\infty \le \theta_n(1 + \|r_n\|_\infty)$ with $\theta_n \to 0$ in probability. If $C > 1$, then $\|u'_n\|_\infty \le C\theta_n(1 + \|r_n\|_\infty)$ with $C\theta_n \to 0$ in probability.

All the requirements of the proposition are satisfied for $(r'_n)_n$ with respect to the ℓ_∞ norm. From our previous step, there exists some $M > 0$ such that $\|r'_n\|_\infty \le M$ with probability one. This implies that $\|r_n\|_\xi \le M$ with probability one and the ℓ_∞ boundedness of $(r_n)_n$ follows from the equivalency of the two norms. ■

3.2.1 *Stochastic approximation for pseudo-contractions*

We are now ready to state and prove the first of our two main convergence results. Instead of proving it for a single pseudo-contraction H, we consider the more general case where we have a sequence of pseudo-contractions $(H_n)_n$, with the same fixed point x^*, the same constant β, and with respect to the same weighted supremum norm.

Proposition 3.2.15: *Let $(r_n)_n$ be the sequence generated by the iteration*

$$r_{n+1} = (1-\gamma_n)r_n + \gamma_n(H_n r_n + w_n + u_n), \tag{3.2.33}$$

where

(a) the step sizes $(\gamma_n)_n$ are such that $\sum_{n=1}^{\infty}\gamma_n(i) = \infty$ and $\sum_{n=1}^{\infty}\gamma_n(i)^2 < \infty$ for every $i = 1,\dots,N$.

(b) The noise terms $(w_n)_n$ have the properties that

$$\mathbb{E}[w_n(i) \mid \mathcal{F}_n] = 0 \quad \text{and} \quad \mathbb{E}\left[w_n(i)^2 \mid \mathcal{F}_n\right] \le A + B\|r_n\|^2.$$

(c) Each H_n is a pseudo-contraction (see Definition 3.2.5) with respect with the same $\|\cdot\|_\xi$ norm, with the same fixed point r^* and the same constant $\beta \in [0,1)$.

(d) There exists a sequence of non-negative random variables $(\theta_n)_n$ which converges to zero almost surely, such that

$$\|u_n\|_\infty \le \theta_n\left(1 + \|r_n\|_\xi\right)$$

for every $n \in \mathbb{N}$.

Then $(r_n)_n$ converges to r^* almost surely.

Proof. Suppose first that $r^* = 0$ and $\xi = (1,\dots,1)$. Then, the pseudo-contraction property can be written as $\|H_n r_n\|_\infty \le \beta\|r_n\|_\infty$ for all n, so Proposition 3.2.14 applies and yields that $(r_n)_n$ is bounded almost surely. Let $D_0 > 0$ be such that $\|r_n\|_\infty \le D_0$ almost surely for every $n \in \mathbb{N}$ and pick some $\varepsilon > 0$ such that $\beta + 2\varepsilon < 1$. We set $D_k = (\beta + 2\varepsilon)^k D_0$ for $k \in \mathbb{N}$.

We will show inductively that for every k, there exists some $n_k \in \mathbb{N}$ such that $\|r_n\|_\infty \le D_k$ for every $n \ge n_k$ almost surely. For $k = 0$, this relation becomes $\|r_n\| \le D_0$ which holds for every $n \in \mathbb{N}$. Suppose that $\|r_n\|_\infty \le D_k$ for every $n \ge n_k$ a.s. for some k.

We define the following sequence $(W_n)_n$ in $\mathbb{R}^N$:

$$W_0 = 0, \quad W_{n+1}(1-\gamma_n)W_n + \gamma_n w_n. \tag{3.2.34}$$

By assupmtion (b), $\mathbb{E}\left[w_n(i)^2 \mid \mathcal{F}_n\right] \le A + B\|r_n\|_2^2$, and since $(r_n)_n$ is bounded a.s., Corollary 3.2.13 applies and asserts that $(W_n)_n$ converges a.s. to zero.

For every $\nu \in \mathbb{N}$, we define

$$W_{\nu,\nu} = 0 \quad \text{and} \quad W_{n+1,\nu} = (1-\gamma_n)W_{n,\nu} + \gamma_n w_n, \quad \text{for } n \ge \nu. \tag{3.2.35}$$

Again by Corollary 3.2.13, $W_{n,\nu} \xrightarrow[n\to\infty]{} 0$ a.s. for every $\nu \in \mathbb{N}$. In addition, for every $n \ge n_k$,

$$\|u_n\|_\infty \le \theta_n\left(1 + \|r_n\|_\xi\right) \le \theta_n(D_k + 1) \xrightarrow[n\to\infty]{} 0 \text{ a.s.}$$

We pick an $\nu_k \geq n_k$ such that $\|u_n\|_\infty \leq \varepsilon D_k$ for every $n \geq \nu_k$, and define $Y_{\nu_k} = (D_k, \dots, D_k)$, and

$$\begin{aligned} Y_{n+1} &= (1-\gamma_n) Y_n + \gamma_n(\beta D_k + \varepsilon D_k) \\ &= Y_n + \gamma_n(\beta D_k + \varepsilon D_k - Y_n) \quad \text{for } n \geq n_k. \end{aligned} \tag{3.2.36}$$

We will show that

$$-Y_n + W_{n,\nu_k} \leq r_n \leq Y_n + W_{n,\nu_k} \tag{3.2.37}$$

for every $n \geq \nu_k$ using induction: For $n = \nu_k$, the inequality becomes $-D_k \leq r_n \leq D_k$ for all $n \geq \nu_k$ which follows from our previous inductive assumption. Suppose that (3.2.37) holds for som $n \geq \nu_k$. Since $\|H_n r_n\| \leq \beta \|r_n\| \leq \beta D_k$, we have that

$$\begin{aligned} r_{n+1} &= (1-\gamma_n) r_n + \gamma_n(H_n r_n + w_n + u_n) \\ &\leq (1-\gamma_n)(Y_n + W_{n,\nu_k}) + \gamma_n(\beta D_k + w_n + \varepsilon D_k) \\ &= (1-\gamma_n) Y_n + \gamma_n(\beta D_k + \varepsilon D_k) + (1-\gamma_n) W_{n,\nu_k} + \gamma_n w_n \\ &= Y_{n+1} + W_{n+1,\nu_k}. \end{aligned}$$

In the same manner we can prove the other half of (3.2.37) for $n := n+1$.

By summing (3.2.36) from $n = 0$ to $n = m-1$ we obtain

$$Y_m = Y_0 + \sum_{n=0}^{m-1} \gamma_n(\beta D_k + \varepsilon D_k - Y_n). \tag{3.2.38}$$

Notice that by its construction, $Y_{n+1} \in [Y_n, (\varepsilon+\beta)D_k]$ for all n, so $(Y_n)_n$ is increasing and bounded, thus convergent. If we take limits in (3.2.38), we have that $\lim_m Y_m = Y_0 + \sum_{n=0}^{\infty} \gamma_n(\beta D_k + \varepsilon D_k - Y_n)$, and since $\sum_n \gamma_n = +\infty$, it has to be that $\lim_n Y_n = \beta D_k + \varepsilon D_k$. Then, (3.2.37) yields that $\limsup_n \|r_n\| \leq (\beta+\varepsilon) D_k = D_{k+1}$, as promised.

We showed that for every k, there exists some $n_k \in \mathbb{N}$ and some Ω_k of measure one, such that $\|r_n(\omega)\|_\infty \leq D_k$ for every $n \geq n_k$ and every $\omega \in \Omega_k$. By setting $\Omega' = \cap_k \Omega_k$, we have that $P(\Omega') = 1$ and $\|r_n(\omega)\|_\infty \leq D_k$ for every $n \geq n_k$ and every $\omega \in \Omega'$. Since $D_k \to 0$, we obtain that $\|r_n\|_\infty \to 0$ almost surely.

We proved the proposition in the special case where $r^* = 0$ and $\xi = (1, \dots, 1)$. The proof for the general case is based on the same argument as in Proposition 3.2.14, p. 66, and we will only sketch it here to avoid repetition.

Suppose that the proposition holds whenever $(H_n)_n$ is a sequence of pseudo-contractions with $r^* = 0$ and ξ arbitrary. We define $H'_n(x) = H_n(x + r^*) - H(r^*)$ and $r'_n = r_n - r^*$. Then each H'_n is a pseudo-contraction with zero as its fixed point[15] and

$$r'_{n+1} = (1-\gamma_n) r'_n + \gamma_n \left(H'_n r'_n + w_n + u_n\right),$$

[15] As was proven in Remark 3.2.8.

so the proposition applies for $(r'_n)_n$ and $r'_n = r_n - r^* \to 0$ almost surely.

Suppose that the proposition holds whenever $(H_n)_n$ is a sequence of pseudo-contractions with r^* arbitrary and $\xi = (1, \dots, 1)$. Using the isometries S, L, we again define $H'_n = SH_nL$, $r'_n = Sr_n$, $w'_n = Sw_n$ and $u'_n = Su_n$ for all n. It is easy to confirm that

$$r'_{n+1} = (1-\gamma_n)r'_n + \gamma_n\left(H'_n r'_n + w'_n + u'_n\right),$$

and that all the assumptions of the proposition, concerning $(w'_n)_n$ and $(u'_n)_n$, are satisfied.

For the general case, where both r^* and ξ associated with H_n are arbitrary, one needs to consecutively apply the previous two transformations. More specifically, r''_n will be equal to $r''_n = S(r_n - r^*)$ and the proposition applied on it will yield that $r''_n \to 0$ a.s., so $Sr_n \to Sr^*$ and $r_n = LSr_r \to LSr^* = r^*$ almost surely. ■

3.2.2 *Stochastic approximation for monotone operators*

For our last convergence result, we drop the pseudo-contraction assumption, and assume that H is a monotone operator instead. Monotonicity is assumed with respect to the usual pointwise order of $\mathbb{R}^N$, but there also exist results involving more general linear orders [Wai19b].

Proposition 3.2.16: *Let $(r_n)_n$ be the sequence generated by the iteration*

$$r_{n+1} = (1-\gamma_n)r_n + \gamma_n(Hr_n + w_n), \tag{3.2.39}$$

where

(a) the step sizes $(\gamma_n)_n$ are such that $\sum_{n=1}^{\infty}\gamma_n(i) = \infty$ and $\sum_{n=1}^{\infty}\gamma_n(i)^2 < \infty$ for every $i = 1, \dots, N$.

(b) The noise terms $(w_n)_n$ have the properties that

$$\mathbb{E}[w_n(i) \mid \mathcal{F}_n] = 0 \quad \text{and} \quad \mathbb{E}\left[w_n(i)^2 \mid \mathcal{F}_n\right] \le A + B\|r_n\|^2.$$

(c) For the operator H we have that

(i) it is monotone, meaning that $Hx \le Hy$ for every $x \le y$.

(ii) For every $\lambda > 0$ and $r \in \mathbb{R}^N$, the following inequality holds: $Hr - \lambda e \le H(r - \lambda e) \le H(r + \lambda e) \le Hr + \lambda e$, where $e = (1, \dots, 1)$.

(iii) It has a unique fixed point, $Hr^* = r^*$.

If $(r_n)_n$ is bounded a.s., then $(r_n)_n$ converges to r^* almost surely.

Proof. Suppose that $(r_n)_n$ is bounded a.s., and let $\lambda > 0$ be a random variable with $r^* - \lambda e \le r_n \le r^* + \lambda e$.[16] We define two sequences $(L^n)_n$ and $(U^n)_n$ as follows:

$$L^0 = r^* - \lambda e, \quad L^{k+1} = \frac{L^k + HL^k}{2}, \quad k \ge 0,$$
$$U^0 = r^* + \lambda e, \quad U^{k+1} = \frac{U^k + HU^k}{2}, \quad k \ge 0.$$

We can show inductively that

$$HU^k \le U^{k+1} \le U^k \text{ and} \tag{3.2.40}$$
$$HL^k \ge L^{k+1} \ge L^k \tag{3.2.41}$$

for every $k \in \mathbb{N}$. For $k = 0$, $HU^0 = H(r^* - \lambda e)$ and $U^1 = \frac{r^* + \lambda e + H(r^* + \lambda e)}{2}$ and since $H(r^* - \lambda e) \le H(r^* + \lambda e), r^* + \lambda e$ we also have that $H(r^* - \lambda e)$ is less or equal than any convex combination of them. In the same manner, $r^* + \lambda e, H(r^* + \lambda e) \le r^* + \lambda e$, so $U^1 = \frac{r^* + \lambda e}{2} + \frac{H(r^* + \lambda e)}{2} \le r^* + \lambda e = U^0$.

Suppose that $HU^m \le U^{m+1} \le U^m$ for every $m = 0, 1, \dots, k$. Then

$$U^{k+2} = \frac{U^{k+1} + HU^{k+1}}{2} \le \frac{U^k}{2} + \frac{HU^{k+1}}{2} \le \frac{U^k}{2} + \frac{HU^k}{2} = U^{k+1},$$

since $U^{k+1} \le U^k$ by our inductive hypothesis and H is monotone. It remains to be shown that $HU^{k+1} \le U^{k+2}$, but this follows immediately from the fact that $HU^{k+1} \le HU^k \le U^{k+1}$ and $U^{k+2} \in [HU^{k+1}, U^{k+1}]$. The inequalities involving L^k can be proven similarly.

We will show that the sequences $(U^k)_k$ and $(L^k)_k$ converge to r^* almost surely. For every k, we have that $U^k \le \dots \le U^0$, so for each $i = 1, \dots, N$, the sequence $(U^k(i))_k$ is non-increasing. It is also bounded below by $r^*(i)$, as

$$U^{k+1} = \frac{U^k + HU^k}{2} \ge \frac{r^* + HU^k}{2} \ge \frac{r^* + Hr^*}{2} = r^*.$$

Thus, for every $i = 1, \dots, N$, the sequence $(U^k(i))_k$ converges to its infimum. By taking limits in $U^{k+1}(i) = \frac{U^k(i) + HU^k(i)}{2}$, and taking advantage of the continuity of H, we obtain that its limit U has the property that $U = \frac{U + HU}{2}$, so $U = r^*$ is the unique fixed point of H. We work similarly for $(L^k)_k$.

The two sequences defined above, restrict the behavior of $(r_n)_n$. In particular, we will show that for every $k \in \mathbb{N}$, there exists some $n_k \in \mathbb{N}$ such that $L^k \le r_n \le U^k$ for every $n \ge n_k$. For $k = 0$, the claim is trivial, as $L^0 = r^* - \lambda e \le r_n \le r^* + \lambda e = U^0$ holds for every $n \in \mathbb{N}$.

For the inductive step, suppose that there exist $n_1, \dots, n_k$, such that $L^l \le r_n \le U^l$ for every $n \ge n_l$ for $l = 1, \dots, k$. We define a sequence $(W_n)_n$ as follows:

$$W_0 = 0, \quad W_{n+1} = (1 - \gamma_n) W_n + \gamma_n w_n, \text{ for } n \ge 0,$$

[16] For almost every ω, there exists an $M_\omega > 0$ with the property that $\|r_n(\omega)\|_\infty \le M_\omega$ for all n. Then the inequality holds for $\lambda(\omega) = M_\omega + \|r^*\|_\infty$.

and for every $\nu \in \mathbb{N}$, we also define

$$W_{\nu,\nu} = 0, \quad W_{n+1,\nu} = (1-\gamma_n)W_{n,\nu} + \gamma_n w_n, \quad \text{for } n \geq \nu.$$

By Corollary 3.2.13, $W_n \to 0$ and $W_{n,\nu} \to 0$ almost surely for every $\nu \in \mathbb{N}$. Let $(X_n)_{n=n_k}^{\infty}$ be defined as follows:

$$X_{n_k} = U^k, \quad X_{n+1} = (1-\gamma_n)X_n + \gamma_n HU^k, \quad \text{for} n \geq n_k.$$

We will show that for every $n \geq n_k$,

$$r_n \leq X_n + W_{n,n_k}. \tag{3.2.42}$$

For $n = n_k$, the relation becomes $r_n \leq U^k$, which holds. Suppose that $r_l \leq X_l + W_{l,n_k}$ for $l = 1, \ldots, n$. Then

$$\begin{aligned} r_{n+1} &= (1-\gamma_n)r_n + \gamma_n Hr_n + \gamma_n w_n \\ &\leq (1-\gamma_n)(X_n + W_{n,n_k}) + \gamma_n Hr_n + \gamma_n w_n \\ &\leq (1-\gamma_n)(X_n + W_{n,n_k}) + \gamma_n HU^k + \gamma_n w_n \\ &\leq (1-\gamma_n)X_n + \gamma_n HU^k + (1-\gamma_n)W_{n,n_k} + \gamma_n w_n \\ &= X_{n+1} + W_{n+1,n_k} \end{aligned}$$

and relation (3.2.42) has been established.

For each $k \in \mathbb{N}$, let

$$A_k = \left\{ i : U^k(i) \neq HU^k(i) \right\} \quad \text{and} \quad \delta_k = \frac{1}{4} \min_{i \in A_k} \left\{ U^k(i) - HU^k(i) \right\}.$$

Then $\delta_k \geq 0$ and δ_k is zero if and only if $U^k = HU^k$. If $\delta_k = 0$, then $U^k = r^*$ and $U^m = r^*$ for every $m \geq k$, so $r_n \leq U^m$ for every $n \geq n_{k_1}$ trivially.

Suppose that $\delta_k \neq 0$ for every $k \in \mathbb{N}$. Pick an $n'_k \in \mathbb{N}$ such that $n'_k \geq n_k$, $\prod_{t=n_k}^{n'_k}(1-\gamma_t(i)) \leq \frac{1}{4}$[17] and $W_{n,n_k} \leq \delta_k$[18] all hold. Then $r_n \leq U^{k+1}$ for every $n \geq n'_k$. Indeed, let $i \in \{1, \ldots, N\}$. If $U^{k+1}(i) = U^k(i)$, then $r_n(i) \leq U^k(i) = U^{k+1}(i)$. Suppose that $U^{k+1}(i) < U^k(i)$. Then for every $n \geq n'_k$,

$$\begin{aligned} X_{n+1}(i) &= (1-\gamma_n(i))X_n(i) + \gamma_n(i)HU^k \\ &= (1-\gamma_n(i))[(1-\gamma_{n-1}(i))X_{n-1}(i) + \gamma_{n-1}(i)HU^k(i)] + \gamma_n(i)HU^k(i) \\ &= (1-\gamma_n(i))(1-\gamma_{n-1}(i))X_{n-1}(i) + \left[1-(1-\gamma_n(i))(1-\gamma_{n-1}(i))\right]HU^k(i), \end{aligned}$$

and inductively we obtain that for every $n \geq n'_k$,

$$X_n = \prod_{t=n_k}^{n-1}(1-\gamma_t)U^k + \left(1 - \prod_{t=n_k}^{n-1}(1-\gamma_t)\right)HU^k = \lambda U^k + (1-\lambda)HU^k,$$

[17] By taking products in the elementary inequality $1 - x \leq e^{-x}$, we have that $\prod_{t=n_k}^{\infty}(1-\gamma_t(i)) \leq e^{-\sum_{t=n_k}^{\infty}\gamma_t(i)} = 0$, for every $i = 1, \ldots, N$. So for every i there exists some $\tilde{n}_i$ with the property that $\prod_{t=n_k}^{n}(1-\gamma_t(i)) \leq \frac{1}{4}$ for every $n \geq \tilde{n}_i$. Set $\tilde{n} = \max_i\{\tilde{n}_i\}$.

[18] Pick an $\overline{n} \in \mathbb{N}$ such that $W_{n,n_k} \leq \delta_k$ for every $n \geq \overline{n}$. This is possible due to the a.s. convergence of $(W_{n,n_k})_n$. To find a common index, just set $n'_k = \max\{n_k, \tilde{n}, \overline{n}\}$.

for $\lambda < \frac{1}{4}$ and $U^k \geq HU^k$. Hence,

$$\begin{aligned} X_n(i) &\leq \frac{1}{4}U^k(i) + \frac{3}{4}HU^k(i) \\ &= \frac{1}{2}U^k(i) + \frac{1}{2}HU^k(i) - \frac{1}{4}\left(U^k(i) - HU^k(i)\right) \\ &\leq U^{k+1}(i) - \delta_k \text{ and} \\ r_n(i) &\leq X_n(i) + W_{n,n_k}(i) \leq U^{k+1}(i) - \delta_k + W_{n,n_k}(i) \leq U^{k+1}(i) \end{aligned}$$

for all $n \geq n'_k$. Similarly, we obtain that $L^{k+1}(i) \leq r_n(i)$ for all i and $n \geq n''_k$. By setting $n_{k+1} = \max\{n'_k, n''_k\}$, we complete the inductive step. ■

3.3 *Q-Learning*

Q-Learning is an algorithm proposed by Chris Watkins [Wat89] to provide approximate solutions to dynamic programming problems, when we do not have enough information on the underlying model to approach it with the usual methods. Despite the relative simplicity of the algorithm itself, the actual proof of its convergence came later and is quite involved. In this section we will present a proof, belonging to John Tsitsiklis [Tsi94], that builds upon the tools developed in the previous sections.

3.3.1 *Dynamic Programming*

In Dynamic Programming [Ros83], we are trying to minimize the expected cost of a process, through decisions that are taken sequentially. During each stage n, we observe the current state $i \in S$ of the process and we choose an action $a \in A(i)$. Then, the process transitions to a state j according to some probability distribution $p_{i\cdot}(a)$ which depends on the previous state i and the decision a that we took. Additionally, the transition we just described, incurs a cost of $c(i, a, j)$. The costs accumulate over time, and our goal is to minimize the total expected cost.

A *policy* π is a sequence $\pi = (\mu_0, \mu_1, \ldots)$, where each μ_k is a function $\mu_k : S \to \cup_{i \in S} A(i)$ with $\mu_k(i) \in A(i)$ for all $i \in S$. Namely, a policy dictates which action we will choose at each state and stage of the problem. A *stationary policy* $\pi = (\mu, \mu, \ldots) = \mu$, is a policy which is indifferent to the current stage of the problem, and depends only on the current state.

Suppose that we follow a policy $\pi = (\mu_0, \mu_1, \ldots)$ and that the sequence of the states the process visits is $(j_k)_0^\infty$. This sequence is not known in advance, but each j_{k+1} depends on the previous state j_k and the decision that was taken during the k-th round. However, this decision was dictated by μ_k and, in particular, it was equal to $\mu_k(i_k)$. So, the total expected cost $J^\pi(i)$ when the initial state is $i_0 = i$ and policy π is employed, is equal to

$$J^\pi(i) = \lim_{N \to \infty} \mathbb{E}\left[\sum_{k=0}^{N} c\left(i_k, \mu_k(i_k), i_{k+1}\right) \mid i_0 = i \right], \tag{3.3.1}$$

provided that this limit exists. The *optimal cost-to-go function* $J^*(i)$ when starting at state i, is defined as the best we can do, under any policy π:

$$J^*(i) = \min_\pi J^\pi(i). \tag{3.3.2}$$

Perhaps the most important property of J^*, is that it satisfies a functional equation called *Bellman's equation:*

$$J^*(i) = \min_{a \in A(i)} \left\{ \sum_{j \in S} p_{ij}(a) \left(c(i, a, j) + J^*(j) \right) \right\}. \tag{3.3.3}$$

This equation also demonstrates the difficulty when one is trying to solve a dynamic programming problem. For example, the rather naive strategy of minimizing the cost in just one step forward,[19] may not be an optimal strategy, because it may cause the process to visit states which will lead to large incurred costs in later rounds. Bellman's equation suggests that in order to solve the problem, one needs to take into an account both the expected cost in one step, $\sum_{j\in S} p_{ij}(a)c(i,a,j)$ and in the later rounds, $\sum_{j\in S} p_{ij}(a)J^*(j)$.

For simple problems, solving the Bellman equation is possible, but in general it is difficult to find algebraic solutions of it. An approximate method of solving it, is the *value iteration method*, according to which we initialize with some function J_0, and at each stage $n+1$ we define

$$J_{n+1}(i) = \min_{a\in A(i)} \left\{ \sum_{j\in S} p_{ij}(a)\left(c(i,a,j) + J_n(j)\right) \right\}. \tag{3.3.4}$$

The Value Iteration Algorithm	
Step 1	Initialize with some function $J_0 : S \to S$.
Step 2	Suppose that J_n has been constructed. During stage $n+1$, set $J_{n+1}(i) = \min_{a\in A(i)} \left\{\sum_{j\in S} p_{ij}(a)\left(c(i,a,j) + J_n(j)\right)\right\}$.
Step 3	Iterate, by returning to Step 2.

Table 3: The Value Iteration Algorithm.

The resulting sequence of functions $(J_n)_n$ converges to J^* uniformly, regardless of the initial choice of J_0. This is a consequence of the contractivity of the operator $T : C(S) \to C(S)$, defined as

$$(Tf)(i) := \min_{a\in A(i)} \left\{ \sum_{j\in S} p_{ij}(a)\left(c(i,a,j) + f(j)\right) \right\}, \ f \in C(S), \ i \in S,$$

and Banach's fixed point theorem.

In the usual dynamic programming setting, all the parameters involved in the problem (incurred costs, probability distributions, decisions and process states) are known in advance to the player. Due to the non-linearity of the Bellman equation, even in this setting, one usually has to settle with approximate solutions. But what happens if we go one step further and assume no prior knowledge on the underlying model? Is it possible to modify the value iteration method in a way to still be able to approximate a solution through it?

More specifically, we consider the same dynamic programming setting as before, with the following modifications: (i) We do not know what the probability

[19] Often called *greedy* or *myopic* strategy.

distributions $p_{i\cdot}(a)$ are, for any $i \in S$ or $a \in A(i)$. (ii) We do not know what the cost functions c are. To compensate with this lack of information, we assume that we can simulate the process, meaning that we can draw values from the unknown distributions $p_{i\cdot}(a)$. We also assume that we can observe the incurred costs after any such simulation.

The Q-Learning algorithm comes to provide a solution for a specific class of problems called *stochastic shortest path problems.*

3.3.1.1 *Stochastic shortest path problems*

In this paragraph we describe a special class of dynamic programming problems, governed by the following assumptions: There exists an absorbing state, denoted as $i = 0$, and when the system visits it, it remains there with no further cost. We also assume that termination is inevitable, and the goal is to minimize the expected cost until termination. In our treatment, the state space $S = \{0, 1, \dots, n\}$ as well as the action sets $A(i)$ at each $i \in S$, are all assumed to be finite.

We will say that a stationary policy is *proper,* if when using this policy, there is a positive probability that the termination state will be reached after at most n stages, regardless of the initial state i_0. As usual, for any stationary policy μ, we introduce the operators T and $T_\mu : C(S) \to C(S)$. For every $J \in C(S) = \mathbb{R}^{n+1}$,

$$(TJ)(i) = \min_{a \in A(i)} \left\{ \sum_{j=0}^{n} p_{ij}(a) \big(c(i,a,j) + J(j) \big) \right\} \quad \text{and} \tag{3.3.5}$$

$$(T_\mu J)(i) = \sum_{j=0}^{n} p_{ij}(\mu(i)) \big(c(i,\mu(i),j) + J(j) \big) \quad \text{for } i = 0, 1, \dots, n. \tag{3.3.6}$$

The components that correspond to $i = 0$ are all equal to zero and will be omitted from now on.

Proposition 3.3.1: [BT96, Proposition 2.1]. *Consider a stochastic shortest path problem for which there exists at least one proper policy and such that for every improper policy μ, there exists some i with $J_\mu(i)$ infinite. Then*

(a) The optimal cost-to-go function J^* has all of its components finite and is the unique fixed point of T, that is, $TJ^* = J^*$.

(b) We have that $\lim_{k\to\infty} T^k J = J^*$ for every $J \in C(S)$.

(c) A stationary policy μ is optimal if and only if $T_\mu J^* = TJ^*$.

(d) For every proper policy μ, its value J^μ is the unique fixed point of T_μ, and additionally, $\lim_{k\to\infty} T_\mu^k J = J^\mu$ for every J.

The operators T and T_μ are also contractions with respect to some appropriate weighted supremum norm.[20]

[20] See also the discussion that follows Definition 3.2.3 for more details on these norms.

Definition 3.3.2: Let $\xi \in \mathbb{R}^N$ with $\xi_i > 0$ for every $i = 1, \dots, N$. The function $\|\cdot\|_\xi : \mathbb{R}^N \to \mathbb{R}$ defined as

$$\|x\|_\xi = \max_{i=1,\dots,N} \frac{|x_i|}{\xi_i} \tag{3.3.7}$$

for $x \in \mathbb{R}^N$ is called the *weighted supremum norm induced by* ξ.

Proposition 3.3.3: *Consider a stochastic shortest path problem for which all the stationary policies are proper. Then, there exists a strictly positive vector* ξ *such that the maps* T *and* T_μ *are contractions with respect to the weighted supremum norm* $\|\cdot\|_\xi$ *for all stationary policies* μ*. In particular, there exists some* $\beta < 1$ *such that*

$$\sum_{j=1}^{n} p_{ij}(a)\xi_j \le \beta \xi_i, \quad \text{for every } i = 1, \dots, n \text{ and } a \in A(i). \tag{3.3.8}$$

Proof. We introduce a modified stochastic shortest path problem which has the same states $S = \{0, 1, \dots, n\}$, action sets $A(i)$ and transition probabilities $p_{ij}(a)$ as the original one, except that all the costs from the non-terminal states are now equal to -1, namely $c(i, a, j) = -1$ for every $i = 1, \dots, n$, every $a \in A(i)$ and every $j = 0, 1, \dots, n$.

By letting all such costs be negative, we are in fact giving the player the incentive to continue playing for as long as possible. Additionally, since all states contribute the same to the player's fortune, the player is rather indifferent to which states he visits, as far as the incurred costs are concerned, and his only goal is to visit the ones which will help him prolong the game.[21]

Let $\tilde{J}$ denote the optimal cost-to-go function for the modified problem. Then for every stationary policy μ and every $i = 1, \dots, n$,

$$\begin{aligned} \tilde{J}(i) &= \min_{a \in A(i)} \left\{ \sum_{j=0}^{n} p_{ij}(a) \left(-1 + \tilde{J}(j)\right) \right\} = -1 + \min_{a \in A(i)} \left\{ \sum_{j=1}^{n} p_{ij}(a)\tilde{J}(j) \right\} \\ &\le -1 + \sum_{j=1}^{n} p_{ij}(\mu(i))\tilde{J}(j). \end{aligned} \tag{3.3.9}$$

Let $\xi_i = \tilde{J}(i)$ for $i = 1, \dots, n$ and $\beta = \max_{i=1,\dots,n} \frac{\xi_i - 1}{\xi_i}$. We will show that $\beta \in [0, 1)$.

Since $c(i, a, j) \le 0$ for every i, j and $a \in A(i)$, we also have that

$$\tilde{J}(i) = \lim_{N \to \infty} \left[\sum_{k=0}^{N-1} c(i_k, a_k, i_{k+1}) \mid i_0 = i \right] \le 0,$$

so $\xi_i = -\tilde{J}(i) \ge 0$. Additionally, for every $i \neq 0$,

$$\tilde{J}(i) = -1 + \min_{a \in A(i)} \left\{ \sum_{j=0}^{n} p_{ij}(a)\tilde{J}(j) \right\} \le 1,$$

[21] Note also that since all the stationary policies are proper, the process will reach the terminal state $i = 0$ at some point with probability one, no matter what policy the player chooses. So J_μ will always be finite.

so $\xi_i \geq 1$.

If $p_{i0}(a) = 1$ for every $a \in A(i)$, then the problem is trivial. So, lets assume that there exists some $i_0 \in \{1, \ldots, n\}$ and $a_0 \in A(i_0)$ with $pi_0,0(a_0) < 1$. Then $\tilde{J}(i_0) = -1 + \sum_{j=0}^{n} p_{i_0 j}(a_0^*)\tilde{J}(j)$ for some $a_0^* \in A(i_0)$. If $p_{i_0,0}(a_1) = 1$ for some $a_1 \in A(i_0)$, then $-1 + \sum_{j=0}^{n} p_{i_0 j}(a_1)\tilde{J}(j) = -1$, whereas

$$\begin{aligned}
-1 + \sum_{j=0}^{n} p_{i_0 j}(a_0)\tilde{J}(j) &= -1 + \sum_{j=1}^{n} p_{i_0 j}(a_0)\tilde{J}(j) \\
&\leq -1 - \sum_{j=1}^{n} p_{i_0 j}(a_0) \\
&= -1 - (1 - p_{i_0 0}(a_0)) \\
&< -1 \\
&= -1 + \sum_{j=0}^{n} p_{i_0 j}(a_1)\tilde{J}(j),
\end{aligned}$$

so a decision which leads to the terminal state with probability one, is never optimal. Consequently, $\tilde{J}(i_0) < -1$ and $\xi_{i_0} > 1$. Since $\xi_i \geq 0$ for all i and there exists some i_0 with $\xi_{i_0} - 1 > 0$, the quantity $\beta = \max_{i=1,\ldots,n} \frac{\xi_i - 1}{\xi_i}$ belongs to $[0, 1)$ as promised.[22]

Let $J, J' \in C(S)$. Then for every stationary policy μ,

$$\begin{aligned}
\left|T_\mu J(i) - T_\mu J'(i)\right| &= \Big| \sum_{j=0}^{n} p_{ij}(\mu(i)) \left(c(i, \mu(i), j) + J(j)\right) \\
&\qquad - \sum_{j=0}^{n} p_{ij}(\mu(i)) \left(c(i, \mu(i), j) + J'(j)\right) \Big| \\
&= \sum_{j=0}^{n} p_{ij}(\mu(i)) \cdot \left|J(j) - J'(j)\right| \\
&\leq \sum_{j=1}^{n} p_{ij}(\mu(i)) \cdot |J(j) - J'(j)| \\
&= \sum_{j=1}^{n} p_{ij}(\mu(i)) \cdot \xi_j \cdot \frac{|J(j) - J'(j)|}{\xi_j} \\
&\leq \sum_{j=1}^{n} p_{ij}(\mu(i)) \xi_j \left\|J - J'\right\|_\xi \\
&\leq \beta \xi_i \left\|J - J'\right\|_\xi . \qquad (3.3.10)
\end{aligned}$$

By dividing with ξ_i and taking suprema, we obtain that T_μ is a $\|\cdot\|_\xi$ contraction. Concerning the operator T, by (3.3.10) we have that

$$T_\mu J(i) \leq T_\mu J'(i) + \beta \xi_i \left\|J - J'\right\|_\xi$$

[22] Also for all stationary policies μ and states i, we have that $\sum_{j=1}^{n} p_{ij}(\mu(i))\xi_j \leq \xi_i - 1 \leq \beta \xi_i$, so the "in particular" part of the proposition has been established.

for every $i = 1, \dots, n$ and every stationary policy μ. So

$$\begin{aligned}\min_\mu T_\mu J(i) =: TJ(i) &\le \min_\mu \left\{ T_\mu J'(i) + \beta \xi_i \left\| J - J' \right\|_\xi \right\} \\ &= \min_\mu T_\mu J'(i) + \beta \xi_i \left\| J - J' \right\|_\xi \\ &= TJ'(i) + \beta \xi_i \left\| J - J' \right\|_\xi ,\end{aligned}$$

and $TJ(i) - TJ'(i) \le \beta \xi_i \left\| J - J' \right\|_\xi$. By reversing the role of J and J' we obtain that $|TJ(i) - TJ'(i)| \le \beta \xi_i \left\| J - J' \right\|_\xi$ for all i, so once again, dividing with ξ_i and taking suprema yields the desired result. ■

3.3.2 *The Q-Learning algorithm and its convergence*

We consider a stochastic shortest path problem with state space $S = \{0, 1, \dots, N\}$, action sets $A(i)$ and costs $c(i, a, j)$ for $i = 1, \dots, N$, $a \in A(i)$ and $j \in S$. For every $(i, a) \in S \times A(i)$, we define the *optimal Q-factor* $Q^*(i, a)$ as

$$Q^*(0, a) = 0 \text{ and } Q^*(i, a) = \sum_{j=0}^{N} p_{ij}(a) \left(c(i, a, j) + J^*(j) \right) \text{ for } i = 1, \dots, N, \tag{3.3.11}$$

where J^* is the optimal cost-to-go function, satisfying the Bellman equation,

$$J^*(j) = \min_{a \in A(i)} \left\{ \sum_{j=0}^{N} p_{ij}(a) \left(c(i, a, j) + J^*(j) \right) \right\}. \tag{3.3.12}$$

Plugging Bellman's equation into (3.3.11), we obtain that

$$Q^*(i, a) = \sum_{j=0}^{N} p_{ij}(a) \left(c(i, a, j) + \min_{b \in A(j)} Q^*(j, b) \right) \text{ for } i = 1, \dots, N, \tag{3.3.13}$$

so the optimal Q-factors satisfy the functional equation (3.3.13). In fact, they are the unique solutions of it:

Proposition 3.3.4: *Consider a stochastic shortest path problem and suppose that J satisfies (3.3.13) and also that $J(0, a) = 0$ for all a. Then J is the optimal Q-factor Q^*.*

Proof. Set $\tilde{J}(i) = \min_{a \in A(i)} J(i, a)$ for $i = 0, \dots, N$. In view of (3.3.13), $\tilde{J}$ satisfies

$$\begin{aligned}\tilde{J}(i) = \min_{a \in A(i)} J(i, a) &= \min_{a \in A(i)} \sum_{j=0}^{N} p_{ij}(a) \left(c(i, a, j) + \min_{b \in A(j)} J(j, b) \right) \\ &= \min_{a \in A(i)} \sum_{j=0}^{N} p_{ij}(a) \left(c(i, a, j) + \tilde{J}(j) \right),\end{aligned}$$

which is just the Bellman equation. By the uniqueness of its solution, we have that $\tilde{J} = J^*$, which plugged into (3.3.13) yields that

$$\begin{aligned} J(i,a) &= \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \tilde{J}(j)\right) \\ &= \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + J^*(j)\right) \\ &=: Q^*(i,a) \end{aligned}$$

for all $i = 1, \ldots, N$ and $a \in A(i)$. For $i = 0$, we also have that $J(0,a) = Q^*(0,a) = 0$, so $J = Q^*$. ■

A direct analogue of the value iteration algorithm applied to Q-factors, would be the iteration

$$Q(i,a) := \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b \in A(j)} Q(j,b)\right) \tag{3.3.14}$$

In the spirit of the Robbins-Monro approximation, a more general version of it, is to consider the convex combination

$$Q(i,a) := (1-\gamma)Q(i,a) + \gamma \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b \in A(j)} Q(j,b)\right) \tag{3.3.15}$$

of the previous value of $Q(i,a)$ with the newly proposed one, with a weight of $\gamma \in (0,1]$. However, in the absence of any knowledge regarding the probability distributions, or even the cost functions, it is not possible to solve (3.3.15) with the standard methods.

Q-Learning [Wat89] is an algorithm that addresses this problem. Although the transition distributions and the cost functions are not known in advance, we assume that we can simulate values from them. In particular, we assume that at each state (i,a) we can generate a new state j according to the transition distribution $p_{i\cdot}(a)$ and observe a cost of $c(i,a,j)$.

The algorithm is essentially a combination of the Robbins-Monro stochastic approximation and of the value iteration methods. In the absence of any knowledge of the transition probabilities $p_{ij}(a)$, the expectation appearing in the right hand side of (3.3.15) is replaced by a single value j which is drawn by the distribution $p_{i,\cdot}(a)$. Similarly, $c(i,a,j)$ is the observed cost when the aforementioned event occurred. The stepsizes γ are also allowed to vary with each iteration n and to depend on the current state (i,a). Most commonly, they are chosen so that $\sum_{n=0}^{\infty} \gamma_n(i,a) = \infty$ and $\sum_{n=0}^{\infty} \gamma_n(i,a)^2 < \infty$ for every i and $a \in A(i)$.

$$Q_{n+1}(i,a) = \left(1 - \gamma_n(i,a)\right) Q_n(i,a) + \gamma_n(i,a)\left(c(i,a,j) + \min_{b \in A(j)} Q_n(j,b)\right),$$

The sequence $(Q_n(i,a))_n$ defined by the previous iteration, converges almost surely to the optimal Q-factor $Q^*(i,a)$ for every $i \in S$ and $a \in A(i)$. The proof we present here belongs to John Tsitsiklis [Tsi94].

The Q-Learning Algorithm	
Step 1	Pick a sequence of functions $(\gamma_n)_n$ defined on the set $\tilde{S} = \{(i,a) : i = 1,\dots,N, a \in A(i)\}$, with the property that $\sum_{n=0}^{\infty} \gamma_n(i,a) = \infty$ and $\sum_{n=0}^{\infty} \gamma_n(i,a)^2 < \infty$ for every $i = 1,\dots,N$ and $a \in A(i)$.
Step 2	Initialize with some function $Q_0(i,a)$ for $i = 1,\dots,N$ and $a \in A(i)$.
Step 3	Suppose that $Q_n(i,a)$ has been generated for all such (i,a) for some $n \geq 0$. For every (i,a), draw a value $j_{i,a}$ according to the distribution $p_{i,\cdot}(a)$ and set $Q_{n+1}(i,a) = (1-\gamma_n(i,a))Q_n(i,a) + \gamma_n(i,a)\left(c(i,a,j_{i,a}) + \min_{b\in A(j_{i,a})} Q_n(j_{i,a},b)\right)$.
Step 4	Return to Step 3.

Table 4: The Q-Learning Algorithm.

Theorem 3.3.5: [BT96, Proposition 5.5. (a)] *Consider the Q-Learning iteration*

$$Q_{n+1}(i,a) = \big(1-\gamma_n(i,a)\big)\,Q_n(i,a) + \gamma_n(i,a)\left(c(i,a,j) + \min_{b\in A(j)} Q_n(j,b)\right), \quad (3.3.16)$$

where j has been drawn according to the distribution $p_{i,\cdot}(a)$ and $(\gamma_n)_n$ is such that $\sum_{n=0}^{\infty}\gamma_n(i,a) = \infty$ and $\sum_{n=0}^{\infty}\gamma_n(i,a)^2 < \infty$ for every $i = 1,\dots,N$ and $a \in A(i)$. If all policies are proper, then $Q_n(i,a) \to Q^*(i,a)$ for every i, $a \in A(i)$ almost surely, where Q^* is the optimal Q-factor.

Proof. Let $\tilde{S} = \{(i,a) : i = 1,\dots,N, a \in A(i)\}$. This is a finite set and each Q-vetor Q is a function $Q : \tilde{S} \to \mathbb{R}$ and can be viewed either as an element in $C(\tilde{S})$ or as a vector in $\mathbb{R}^{|\tilde{S}|}$. We define the operator $H : C(\tilde{S}) \to C(\tilde{S})$ as follows. For every Q in $C(\tilde{S})$,

$$(HQ)(i,a) = \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b\in A(j)} Q(j,b)\right). \quad (3.3.17)$$

By letting w_n be equal to

$$w_n(i,a) = c(i,a,j) + \min_{b\in A(j)} Q_n(j,b) - \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b\in A(j)} Q_n(j,a)\right), \quad (3.3.18)$$

the Q-Learning iteration (3.3.16) can be re-written as

$$Q_{n+1}(i,a) = (1-\gamma_n(i,a)) + Q_n(i,a) + \gamma_n(i,a)\,(HQ_n(i,a) + w_n(i,a)), \quad (3.3.19)$$

with

$$\mathbb{E}[w_n(i,a)\,|\,\mathcal{F}_n] = \sum_{j=0}^{N} p_{ij}(a)c(i,a,j) + \sum_{j=0}^{N} p_{ij}(a) \min_{a\in A(j)} Q_n(j,a) -$$

$$
\begin{aligned}
&\quad - \sum_{j=0}^{N} p_{ij}(a)c(i,a,j) - \sum_{j=0}^{N} p_{ij}(a) \min_{b\in A(j)} Q_n(j,a) \\
&= 0, \text{ and} \\
\mathbb{E}\left[w_n(i,a)^2 \mid \mathcal{F}_n\right] &= \mathbb{E}\left[c(i,a,j)^2 + \min_{a\in A(j)} Q_n(j,a)^2 \,\middle|\, \mathcal{F}_n\right] + \\
&\quad + \left[\sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b\in A(j)} Q_n(j,a)\right)\right]^2 + \\
&\quad + \mathbb{E}\left[2c(i,a,j) \min_{a\in A(j)} Q_n(j,a) \,\middle|\, \mathcal{F}_n\right] - \\
&\quad - 2\sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b\in A(j)} Q_n(j,a)\right)\mathbb{E}[c(i,a,j) \mid \mathcal{F}_n] - \\
&\quad - 2\sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b\in A(j)} Q_n(j,a)\right)\mathbb{E}\left[\min_{a\in A(j)} Q_n(j,a)^2 \,\middle|\, \mathcal{F}_n\right] \\
&\leq K^2 + \max_{j,a} Q_n^2(j,a) + 2K \max_{j,a} Q_n(j,a)^2 + \left(K + \max_{j,a} Q_n(j,a)^2\right)^2 \\
&\leq \Lambda\left(1 + \max_{j,a} Q(j,a)^2\right),
\end{aligned}
$$

where $K = \max\{c(i,a,j) : i = 1,\dots,n, a \in A(i)\}$ and $\Lambda = 2K^2+4K+2$. So the noise terms $(w_n)_n$, satisfy the assumptions of Proposition 3.2.15.

By Proposition 3.3.3, there exists a strictly positive $\xi \in \mathbb{R}^N$ such that

$$\sum_{j=1}^{N} p_{ij}(a)\xi_j \leq \beta\xi_i$$

for every $i = 1,\dots,N$ and $a \in A(i)$. Each Q can be seen as a vector in $\mathbb{R}^M$, where M is the cardinality of the set $\tilde{S}$. In particular, if one sets $k_i = |A(i)|$, then $Q : \mathbb{R}^{k_1+\dots+k_N} \to \mathbb{R}$. We extend $\xi \in \mathbb{R}^N$ into a vector $\tilde{\xi} \in \mathbb{R}^{k_1+\dots+k_N}$, by just repeating each coordinate i of ξ, a total of k_i times:

$$\tilde{\xi}_l = \begin{cases} \xi_1, & l \in [1, k_1], \\ \xi_2, & l \in (k_1, k_1+k_2], \\ \vdots & \vdots \\ \xi_n, & l \in \left(\sum_{i=1}^{n-1} k_i, \sum_{i=1}^{n} k_i\right]. \end{cases}$$

We can define the $\tilde{\xi}$ weighted supremum norm on $\mathbb{R}^{k_1+\dots+k_N}$, and by the definition of ξ it has the property that

$$\|Q\|_{\tilde{\xi}} = \max_{i,a\in A(i)} \frac{|Q(i,a)|}{\xi_i} \tag{3.3.20}$$

for every $Q \in \mathbb{R}^{k_1+\dots+k_N}$. Let $Q, Q' \in \mathbb{R}^{k_1+\dots+k_N}$. Then

$$\left|HQ(i,a) - HQ'(i,a)\right| = \left|\sum_{j=1}^{N} p_{ij}(a)\left(\min_{a\in A(j)} Q(j,a) - \min_{a\in A(j)} Q'(j,a)\right)\right|$$

$$
\begin{aligned}
&\leq \sum_{j=1}^{N} p_{ij}(a) \left| \min_{a \in A(j)} Q(j,a) - \min_{a \in A(j)} Q'(j,a) \right| \\
&\leq \sum_{j=1}^{N} p_{ij}(a) \max_{j, a \in A(j)} \left| Q(j,a) - Q'(j,a) \right| \\
&= \sum_{j=1}^{N} p_{ij}(a) \xi_j \cdot \max_{j, a \in A(j)} \frac{|Q(j,a) - Q'(j,a)|}{\xi_j} \\
&= \sum_{j=1}^{N} p_{ij}(a) \xi_j \cdot \|Q - Q'\|_{\tilde{\xi}} \\
&\leq \beta \xi_i \|Q - Q'\|_{\tilde{\xi}},
\end{aligned}
$$

so H is a β-contraction with respect to the weighted supremum norm. By Proposition 3.2.15, $(Q_n)_n$ converges with probability one to the fixed point of H, which is just the optimal Q-factor Q^*. ■

Theorem 3.3.6: [BT96, Proposition 5.5. (b)] *Consider the Q-Learning iteration*

$$
Q_{n+1}(i,a) = (1 - \gamma_n(i,a)) Q_n(i,a) + \gamma_n(i,a) \left(c(i,a,j) + \min_{b \in A(j)} Q_n(j,b) \right), \quad (3.3.21)
$$

where j has been drawn according to the distribution $p_{i,\cdot}(a)$ and $(\gamma_n)_n$ is such that $\sum_{n=0}^{\infty} \gamma_n(i,a) = \infty$ and $\sum_{n=0}^{\infty} \gamma_n(i,a)^2 < \infty$ for every $i = 1, \ldots, N$ and $a \in A(i)$. Suppose that there exists at least one proper policy, and that for every improper one μ, the corresponding value $J^{\mu}(i_0)$ is infinite for some state i_0. If $(Q_n)_n$ is bounded with probability one, then $Q_n(i,a) \to Q^*(i,a)$ for every i, $a \in A(i)$ almost surely, where Q^* is the optimal Q-factor.

Proof. The operator H defined in (3.3.17) is monotone: Suppose that $Q \leq Q'$ in the usual order of $\mathbb{R}^M$, where $M = |\{(i,a) : i = 1, \ldots, N, a \in A(i)\}|$. Then

$$
\begin{aligned}
(HQ)(i,a) &= \sum_{j=0}^{N} p_{ij}(a) \left(c(i,a,j) + \min_{b \in A(j)} Q(j,b) \right) \\
&\leq \sum_{j=0}^{N} p_{ij}(a) \left(c(i,a,j) + \min_{b \in A(j)} Q'(j,b) \right) \\
&= (HQ')(i,a)
\end{aligned}
$$

for every i and $a \in A(i)$. The operator H has also a unique fixed point. Indeed, a fixed point Q of H must satisfy the functional equation

$$
Q(i,a) = \sum_{j=0}^{N} p_{ij}(a) \left(c(i,a,j) + \min_{b \in A(j)} Q(j,b) \right) \quad (3.3.22)
$$

for all $i = 1, \ldots, N$ and $a \in A(i)$, and $Q(0,a) = 0$ for all a. As we showed in Proposition 3.3.4, this equation has a unique solution.

Lastly, for every $\lambda > 0$, by taking into account the monotonicity of H and the fact that $Q - \lambda e \leq Q + \lambda e$ for all Q, we have that

$$\begin{aligned}
(HQ - \lambda e)(i,a) &= \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b \in A(j)} Q(j,b)\right) - \lambda \\
&= \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b \in A(j)} Q(j,b) - \lambda\right) \\
&= \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b \in A(j)} (Q(j,b) - \lambda)\right) \\
&= H(Q - \lambda e) \\
&\leq H(Q + \lambda e) \\
&= \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b \in A(j)} Q(j,b) + \lambda\right) \\
&= \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b \in A(j)} Q(j,b)\right) + \lambda \\
&= (HQ + \lambda e)(i,a)
\end{aligned}$$

for every $i, a \in A(i)$. This shows that $HQ - \lambda e \leq H(Q - \lambda e) \leq H(Q + \lambda e) \leq H(Q) + \lambda e$. By Proposition 3.2.16, we have that $(Q_n)_n$ converges to Q^* almost surely, provided that it is bounded. ■

References

[AT07] C. Aliprantis, R. Tourky, *Cones and Duality*, American Mathematical Society, 2007. ISBN: 978-0-8218-4146-4 Cited on p.

[Arg04] S. Argyros, *Lecture Notes in Functional Analysis* (in greek), Athens, 2004. https://semfe.gr/files/users/795/synarthsiakh_analysh_I_biblio.pdf Cited on p. 22, 35, 36, 54, 121, 122, 123, 125

[Ban22] S. Banach, Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales, *Fundamenta Mathematicae,* **3** (1), pp. 133-181, 1922. DOI: 10.4064/fm-3-1-133-181 Cited on p. 52

[BT96] D. Bertsekas, J. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996. ISBN: 1-886529-10-8 Cited on p. 11, 19, 54, 75, 80, 82

[Nev75] J. Neveu, *Discrete Parameter Martingales*, North-Holland, 1975. ISBN: 978-0-7204-2810-0 Cited on p. 135

[Rob52] H. Robbins, Some aspects of the sequential design of experiments, *Bull. Amer. Math. Soc.* **58** (5), pp. 527-535, 1952. DOI: S0002-9904-1952-09620-8 Cited on p. xiv, 13, 86, 88

[RM51] H. Robbins, S. Monro, A Stochastic Approximation Method, *Ann. Math. Statist.* **22** (3), pp. 400-407, 1951. DOI: 10.1214/aoms/1177729586 Cited on p. xiv, 7, 47, 48, 52

[Ros83] S. Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press, 1983. ISBN: 9781483269092 Cited on p. 73

[Tsi94] J. TSITSIKLIS, Asynchronous Stochastic Approximation and Q-Learning, *Machine Learning* **16** (3), pp. 185–202, 1994. DOI: 10.1023/A:102268912 Cited on p. 54, 73, 79

[Wai19] M. WAINWRIGHT, Stochastic approximation with cone-contractive operators: Sharp ℓ_∞-bounds for Q-learning, 2019. arXiv:1905.06265 Cited on p. 69

[Wat89] C. WATKINS, *Learning from Delayed Rewards*, PhD Thesis, King's College, Cambridge, UK, 1989. http://www.cs.rhul.ac.uk/~chrisw/new_thesis.pdf Cited on p. 11, 73, 79

4

Multi-armed Bandits

Multi-armed bandits is a subfield of Reinforcement Learning which was first formulated during the early '50s through the work of Herbert Robbins. In the last decades the interest towards it has re-emerged, resulting in strong contributions both to its foundations and its practical implications. In Section 4.2 we present the Robbins algorithm, which is a very basic argument that achieves maximum mean reward asymptotically. The question of how fast this reward can be approached, requires some heavy machinery, and is studied in detail in Section 4.3 through the Lai-Robbins arguments. In Sections 4.4 and 4.5 we present the Auer-Bianci-Fischer upper confidence bound algorithms, which simplify some of the Lai-Robbins ideas with a relatively small price in performance.

4.1 *Introduction*

In its simplest form, the problem can be stated as follows: We are given the opportunity to draw samples from two different statistical populations A and B, specified by the distributions F_A and F_B, with finite means a and b respectively. Each time we draw a value x, we get to keep it as a reward. Our goal is to draw a sample $x_1, \dots, x_n$ in a way to maximize the expected sum $S_n = x_1 + \dots + x_n$. In each step, we can decide from which population we will draw the next value, based on some rule which will possibly take into consideration the values sampled up to that point.

By the linearity of the expectation,

$$\mathbb{E}[S_n/n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{k_n}{n}a + \frac{n-k_n}{n}b,$$

where k_n is the number of samples drawn from A during the first n repetitions. This means that $\mathbb{E}[S_n/n]$ is a convex combination of a and b and belongs to the interval $[\min\{a,b\}, \max\{a,b\}]$. Clearly, if we knew the actual values of a and b we could maximize $\mathbb{E}[S_n/n]$ by constantly drawing from the population with the largest mean. However, neither the values of a and b, nor their respective order are known in advance.

In order to approach the problem efficiently, we need to employ simultaneously two contradictory strategies. Firstly, we should draw enough values from both populations in order to estimate the means a and b adequately. This implies that we need to be willing to deliberately choose the suboptimal arm a fair amount of times, so as to make sure that it is indeed suboptimal. Secondly, once we are fairly certain on which arm is the optimal one, we should pick it progressively more often, so as to maximize our expected profit.

The first strategy is called *exploration*, whereas the second one *exploitation*, and the multi-armed bandit setting is one of the most natural and didactic paradigms of the so-called *exploration vs exploitation dilemma/trade off*, namely the problem of finding the correct balance between these two conflicting strategies.

4.2 *The Robbins algorithm*

Let ϕ be a drawing rule from the populations A and B. Here $\phi : \mathbb{N} \to \{A, B\}$ is just a function which at each stage n selects a population to draw from. The value of $\phi(n)$ is usually not predetermined, but it depends on the values of the sample obtained thus far until stage n. We define

$$L_n(A, B, \phi) = \max\{a, b\} - \mathbb{E}\left[\frac{S_n}{n}\right], \tag{4.2.1}$$

a quantity that measures the distance between the expected value of S_n, when following the rule ϕ, and the best possible outcome $\max\{a, b\}$ after n draws. Herbert Robbins [Rob52] showed that there always exists a rule ϕ for which

$$\lim_{n\to\infty} L_n(A, B, \phi) = 0,$$

thus one can always approach this maximum asymptotically.

The rule proposed by Robbins, estimates the true values of a and b progressively and chooses the sampling population accordingly. In order to make sure that its estimates for a and b are accurate, the rule has to draw infinitely many samples from both populations and rely on the Law of Large Numbers to distinguish between a and b. However, drawing too many values from the wrong population could potentially move $\mathbb{E}[S_n/n]$ away from its maximum possible value $\max\{a, b\}$.

Robbins ensures that $\mathbb{E}[S_n/n]$ will be unaffected, by choosing to draw from A and B for predetermined indices J_A and $J_B \subseteq \mathbb{N}$. If J_A and J_B are sparse enough, the drawn values will have no effect on the mean value of S_n. For the rest of the indices $\mathbb{N} \setminus (J_A \cup J_B)$, he allows the rule to draw from the population that seems to be the best up to that point. In this context, the "sparseness" of a set is being measured by its natural density, and by sparse subsets of $\mathbb{N}$ we mean infinite sets with zero natural density (see Appendix C.1 for more details):

Definition 4.2.1: Let $J \subseteq \mathbb{N}$. We define its *natural density* $d(J)$ as

$$d(J) = \lim_{n\to\infty} \frac{\#J \cap \{1, \dots, n\}}{n}, \tag{4.2.2}$$

provided that this limit exists.

An example of two disjoint infinite sets with zero density are $J_A = \{n^2 : n \in \mathbb{N}\}$ and $J_B = \{n^2 + 1 : n \in \mathbb{N}\}$.

For proving Robbins result, it will be useful to study the behavior of the Cesàro averages of a sequence when we can "decompose" it into two subsequences with known Cesàro limits, prescribed on index sets of known natural densities. Suppose that we have two sequences $(a_n)_n$ and $(b_n)_n$ and a set $J \subseteq \mathbb{N}$ such that both

J and J^c are infinite. We can "merge" the two sequences into one sequence $(c_n)_n$ according to the following rule:

For the indices that belong to J, c_n will be equal to the first unassigned term of the sequence $(a_n)_n$, whereas for $n \notin J$, c_n will be equal to the first unassigned term of $(b_n)_n$. In this way, the newly defined sequence $(c_n)_n$ contains both $(a_n)_n$ and $(b_n)_n$ as subsequences, the exact positions of which are fully determined by J and J^c respectively. In mathematical terms, $(c_n)_n$ has the property that $(a_n)_{n\in\mathbb{N}} = (c_n)_{n\in J}$ and $(b_n)_{n\in\mathbb{N}} = (c_n)_{n\in J^c}$.

Definition 4.2.2: Let $(a_n)_n$ and $(b_n)_n$ be two sequences, and $J \subseteq \mathbb{N}$ be a set such that both J and J^c are infinite. We define the *merge of* $(a_n)_n$ *and* $(b_n)_n$ *over* (J, J^c), as the sequence $(c_n)_n$ with the property that $(a_n)_{n\in\mathbb{N}} = (c_n)_{n\in J}$ and $(b_n)_{n\in\mathbb{N}} = (c_n)_{n\in J^c}$.

More explicitly, for every $n \in \mathbb{N}$ set $d_n = \#J \cap \{1, \dots, n\}$. Then $(c_n)_n$ can be defined as

$$c_n = \begin{cases} a_{d_n}, & \text{if } n \in J, \\ b_{n-d_n}, & \text{if } n \in J^c. \end{cases}$$

One can easily extend the previous definition in order to merge more than two sequences.[1] Suppose now that we have two Cesàro summable sequences $(a_n)_n$, $(b_n)_n$ and that $J \subseteq \mathbb{N}$ is an infinite set such that J^c is also infinite with $d(J)$ well defined. If we merge the two sequences over (J, J^c), we intuitively expect that the resulting sequence $(c_n)_n$ will also be Cesàro summable and that its sum will be the weighted average of the two individual Cesàro sums, with the weights being equal to the corresponding densities of the partition sets. The following lemma suggests that the merge operation indeed respects the Cesàro limits and index densities.

Lemma 4.2.3: *Let $(a_n)_n$ and $(b_n)_n$ be two Cesàro summable sequences with respective Cesàro limits a and b, and let $J \subseteq \mathbb{N}$ such that both J and J^c are infinite and $d(J)$ is well defined. Then, the merge of $(a_n)_n$ and $(b_n)_n$ over (J, J^c) is also Cesàro summable, and its Cesàro sum is equal to $c = d(J)a + (1 - d(J))b$.*

Proof. Let $(c_n)_n$ denote their merge. Then by definition,

$$\begin{aligned} \frac{c_1 + \ldots + c_n}{n} &= \frac{a_1 + \ldots + a_{d_n}}{n} + \frac{b_1 + \ldots + b_{n-d_n}}{n} \\ &= \frac{a_1 + \ldots + a_{d_n}}{d_n} \cdot \frac{d_n}{n} + \frac{b_1 + \ldots + b_{n-d_n}}{n - d_n} \cdot \frac{n - d_n}{n}, \end{aligned}$$

with $\lim_{n\to\infty} \frac{d_n}{n} = d(J)$ and $\lim_{n\to\infty} \frac{n-d_n}{n} = 1 - d(J) = d(J^c)$.[2] Additionally, we have that $\lim_{n\to\infty} \frac{a_1+\ldots+a_{d_n}}{d_n} = a$ and $\lim_{n\to\infty} \frac{b_1+\ldots+b_{n-d_n}}{n-d_n} = b$, since both $(d_n)_n$ and $(n - d_n)_n$ tend to infinity as $n \to \infty$. ∎

The Robbins Rule	
Step 1	Choose $J_A, J_B \subseteq \mathbb{N}$ infinite sets with zero density.
Step 2	Suppose that a sample $x_1, \dots, x_{n-1}$ has been drawn. If $n \in J_A$, draw X_n from population A. If $n \in J_B$, draw X_n from B.
Step 3	For $n \notin J_A \cup J_B$, set $a_n = \frac{\sum_{\{i:X_i \sim F_A\}} x_i}{\#\{i:X_i \sim F_A\}}$ and $b_n = \frac{\sum_{\{i:X_i \sim F_B\}} x_i}{\#\{i:X_i \sim F_B\}}$. If $a_n \geq b_n$ draw X_n from population A, else draw it from B.

Table 5: The Robbins Rule [Rob52].

We can now state the Robbins rule and prove the asymptotic properties of $\mathbb{E}[S_n/n]$ under it, using the previous Lemma.

Theorem 4.2.4: *Let A, B be two populations with mean a and b respectively, and let $(X_n)_n$ be a sequence of random variables constructed according to the Robbins rule. Then $\mathbb{E}\left[\frac{S_n}{n}\right] \to \max\{a, b\}$.*

Proof. For each $\omega \in \Omega$, let $X_n(\omega)$ be the n-th observation according to this rule. Each $X_n(\omega)$ is either taken from population A or population B, and regardless of the sampling strategy, the sequence $(X_n(\omega))_n$ contains infinite samples from both populations. Let [3]

$$I_A(\omega) = \{n \in \mathbb{N} : X_n(\omega) \text{ has been taken from } A\} \quad \text{and} \quad I_B(\omega) = \mathbb{N} \setminus I_A(\omega).$$

Although the exact elements of $I_A(\omega)$ and $I_B(\omega)$ are not known in advance, it is certain that they must contain J_A and J_B respectively, as on these two sets the population from which the sample will be taken has been predetermined.

By the Strong Law of Large Numbers,

$$\frac{1}{\#\{1,\dots,k\} \cap I_A(\omega)} \sum_{n \in \{1,\dots,k\} \cap I_A(\omega)} X_n(\omega) \xrightarrow[k\to\infty]{} a \quad \text{and}$$
$$\frac{1}{\#\{1,\dots,k\} \cap I_B(\omega)} \sum_{n \in \{1,\dots,k\} \cap I_B(\omega)} X_n(\omega) \xrightarrow[k\to\infty]{} b,$$

for almost every $\omega \in \Omega$. Suppose that $a < b$. Then for almost every $\omega \in \Omega$, there exists a $n_\omega \in \mathbb{N}$ such that $n \in I_B(\omega)$ for every $n \geq n_\omega$ as long as $n \notin J_A$. Additionally, $(X_n(\omega))_n$ can be viewed as the merge of the two sequences $(X_n(\omega))_{n \in I_A(\omega)}$ and $(X_n(\omega))_{n \in I_B(\omega)}$. By Lemma 4.2.3, the Cesàro averages of the sequence $(X_n(\omega))_n$

[1] Let $(a_n^1)_n, \dots, (a_n^N)_n$ be N sequences and $J_1, \dots, J_N$ a partition of $\mathbb{N}$ such that each J_i is an infinite set. We define the merge of them over the partition $\{J_1, \dots, J_N\}$ as the sequence $(c_n)_n$ with the property that $(c_n)_{n \in J_i} = (a_n^i)_n$ for every $i = 1, \dots, N$.

[2] See Lemma C.1.2 for the reason why d satisfies this measure-like property.

[3] The set $I_A(\omega)$ contains J_A and possibly some values of n for $n \leq n_\omega$. Consequently, it is contained in the union of two sets, $I_A(\omega) \subseteq J_A \cup \{1, \dots, n_\omega\}$, both of which have zero density, the former by our assumption and the latter being finite. By Lemma C.1.2, it follows that $I_A(\omega)$ has zero density as well.

converge to $0 \cdot a + 1 \cdot b = b = \max\{a, b\}$ for almost every ω, since $d(I_A) = 0$. Having [4] established the almost sure convergence of $(\frac{S_n}{n})_n$, we can additionally conclude that $\mathbb{E}\left[\frac{S_n}{n}\right] \to \max\{a, b\}$, using the following argument:

Let $\varepsilon > 0$. By the absolute continuity of the Lebesgue integral (Proposition B.2.1), for each $n \in I_A$, there exists some $\delta_A > 0$ such that $\int_E X_n dP < \varepsilon/2$ for every E with $\lambda(E) < \delta_A$. Similarly, for each $n \in I_B$, there exists some $\delta_B > 0$ such that $\int_E X_n dP < \varepsilon/2$ for every E with $\lambda(E) < \delta_B$. Set $\delta = \min\{\delta_A, \delta_B\}$. Then for every E with $\lambda(E) < \delta$,

$$\begin{aligned}\int_E |S_n/n| dm &\leq \frac{k_n}{n} \int_E |X_A| dm + \frac{n - k_n}{n} \int_E |X_B| dm \\ &\leq \frac{k_n}{n} \frac{\varepsilon}{2} + \frac{n - k_n}{n} \frac{\varepsilon}{2} = \varepsilon,\end{aligned}$$

where k_n is the amount of samples taken from A during the first n draws and X_A, X_B are random variables with distributions F_A and F_B respectively. Additionally, $(S_n/n)_n$ is bounded in L_1 by the triangle inequality:

$$\|S_n/n\|_1 = \|(X_1 + \ldots + X_n)/n\|_1 \leq \frac{1}{n} \sum_{i=1}^{n} \|X_i\|_1 \leq \max\{a, b\}.$$

By Theorem B.2.3, $(S_n/n)_n$ is uniformly integrable and by Theorem B.2.4, we obtain that $\mathbb{E}[S_n/n] \to \max\{a, b\}$. ■

4.3 *Asymptotically optimal lower bounds*

The Robbins Rule achieves the maximum possible mean reward asymptotically, but does not address the question of how fast this reward is approached. For real life problems, where one is not allowed to play the game ad infinitum, but has to stop at some finite time n_0, the Robbins Rule could be completely impractical. The rate of convergence question was addressed and answered in an impressive manner by Tze Leung Lai and Herbert Robbins [LR85] more than 30 years after Robbin's original publication.

They showed, that for reward distributions that satisfy certain assumptions, the regret of any algorithm may not grow slower than $o(\ln n)$. They also proposed an algorithm which achieves this lower bound, thus behaving optimally with respect to its asymptotic rate of convergence.

4.3.1 *The Kullback-Leibler divergence*

The Kullback-Leibler divergence [Joy11], [BLM13, Paragraph 4.8], [Wai19a, Paragraph 3.3.2] is a notion of distance between probability distributions that plays

[4] It is known from Probability Theory, that when a sequence of random variables $(X_n)_n$ converges in L_1 to some random variable X, then it converges to it in probability. However, in general, the converse does not hold, so in order to show that the converse holds for $(S_n/n)_n$, we will have to rely on some specific strong properties of it, namely its uniform integrability (see Appendix B.2).

a central role in the bandit setting, both when trying to find asymptotic lower bounds and when constructing algorithms that attain them.

Definition 4.3.1: Let $(\Omega, \mathcal{A}, P)$ be a probability space and suppose that Q is a probability measure on $\mathcal{A}$ which is absolutely continuous with respect to P. Let $Y = \frac{dQ}{dP}$. We define the *Kullback-Leibler* divergence $I(Q,P)$ of Q with respect to P, as

$$\begin{aligned} I(Q,P) :&= \mathbb{E}_P[Y \ln Y] = \int Y \ln Y dP \\ &= \mathbb{E}_Q[\ln Y] = \int \ln Y dQ. \end{aligned}$$

The first line is the usual definition of the Kullback-Leibler divergence, whereas in the second line we changed the measure with respect to which we integrate, using the following theorem:

Theorem 4.3.2: [AB06, Theorem 13.23] *Let μ, ν be sigma-finite measures on the sigma-algebra $\mathcal{A}$ such that ν is absolutely continuous with respect to μ. Let also $g = \frac{d\nu}{d\mu}$ be the Radon-Nikodym derivative of ν with respect to μ. Then, for every ν-integrable function f, the function fg is μ-integrable and*

$$\int f d\nu = \int fg d\mu. \tag{4.3.1}$$

When both Q and P are absolutely continuous with respect to some measure ν with densities $f(x;\lambda)$ and $f(x;\mu)$ respectively, $I(Q,P)$ can also be written as

$$I(Q,P) = I(\lambda,\mu) = \int f(x;\lambda) \ln \frac{f(x;\lambda)}{f(x;\mu)} d\nu(x). \tag{4.3.2}$$

Using Jensen's inequality on the convex function $\Phi(x) = x \ln x$, it is easy to see that $I(Q,P) \geq 0$. Additionally, $I(Q,P) = 0$ if and only if $P = Q$, however I does not meet the rest of the metric axioms in general,[5] hence the term "divergence" instead of "metric".

4.3.2 *The theoretical lower bound*

Lai and Robbins make certain assumptions on the reward distributions. Firstly, they assume that they are all parametric with densities belonging in the family $(f(x;\theta))_{\theta\in\Theta}$ for some one-dimensional parameter space $\Theta \subseteq \mathbb{R}$. Secondly, the Kullback-Leibler divergence between the members of the family $(f(x;\theta))_{\theta\in\Theta}$ satisfies a certain continuity type property, and thirdly, the set Θ itself is "rich" in elements. We will explain what we mean in detail below.

[5] A glance at (4.3.2) reveals that I need not be symmetric, but even if we consider its "symmetrization" $I'(Q,P) = \frac{1}{2}(I(Q,P) + I(P,Q))$, which makes sense when both P and Q are absolutely continuous with respect to each other, then I' may still fail the triangle inequality.

Assumption 4.3.3: *The family of distributions $f(x;\lambda)$ has the properties that $0 < I(\theta,\lambda) < \infty$ whenever $\mu(\lambda) > \mu(\theta)$ and that for every $\varepsilon > 0$ and every $\theta, \lambda \in \Theta$ such that $\mu(\lambda) > \mu(\theta)$, there exists some $\delta > 0$ such that*

$$|I(\theta,\lambda) - I(\theta,\lambda')| < \varepsilon, \quad \text{whenever} \quad \mu(\lambda) \le \mu(\lambda') \le \mu(\lambda) + \delta. \tag{4.3.3}$$

This assumption is in fact a continuity-type property of I with respect, not to its second variable, but rather to its dependence through $\mu(\lambda)$. Using an argument similar to the proof of the sequential property of the usual continuity, we can reformulate this assumption into a more intuitive one involving sequences.

Proposition 4.3.4: *Let $f(x;\lambda)$ be a family of distributions. The following are equivalent:*

(a) For every $\varepsilon > 0$ and every $\theta, \lambda \in \Theta$ such that $\mu(\lambda) > \mu(\theta)$, there exists some $\delta > 0$ such that $\left|I(\theta,\lambda) - I(\theta,\lambda')\right| < \varepsilon$ whenever $\mu(\lambda) \le \mu(\lambda') \le \mu(\lambda) + \delta$.

(b) For every $\lambda, \theta \in \Theta$ with $\mu(\lambda) > \mu(\theta)$, and every real sequence $(\lambda_n)_n$ such that $\mu(\lambda_n) \downarrow \mu(\lambda)$,[6] we have that $I(\theta,\lambda_n) \to I(\theta,\lambda)$.

Proof. (a) $\Rightarrow$ (b) Fix $\theta, \lambda \in \Theta$ with $\mu(\lambda) > \mu(\theta)$ and pick a sequence $(\lambda_n)_n$ such that $\mu(\lambda_n) \downarrow \mu(\lambda)$. Let $\varepsilon > 0$. There exists a $\delta > 0$ such that for every λ' with $\mu(\lambda) \le \mu(\lambda') \le \mu(\lambda) + \delta$, we have that $\left|I(\theta,\lambda) - I(\theta,\lambda')\right| < \varepsilon$. For this δ, there exists some $n_0 \in \mathbb{N}$ such that $\mu(\lambda) \le \mu(\lambda_n) \le \mu(\lambda) + \delta$ for every $n \ge n_0$, and thus $|I(\theta,\lambda) - I(\theta,\lambda_n)| < \varepsilon$ for every $n \ge n_0$.

(b) $\Rightarrow$ (a) Suppose that (a) does not hold. Then there exists some $\varepsilon > 0$ and some $\theta, \lambda \in \Theta$ with $\mu(\lambda) > \mu(\theta)$, such that for every $\delta > 0$, there exists some λ_δ such that $\mu(\lambda) \le \mu(\lambda_\delta) \le \mu(\lambda) + \delta$ and $|I(\theta,\lambda) - I(\theta,\lambda_\delta)| \ge \varepsilon$.

By setting $\delta = \frac{1}{n}$, we can inductively construct a sequence $(\lambda_n)_n$ such that

$$\mu(\lambda) \le \mu(\lambda_n) \le \mu(\lambda) + \frac{1}{n} \quad \text{and} \quad |I(\theta,\lambda) - I(\theta,\lambda_n)| \ge \varepsilon \tag{4.3.4}$$

for all $n \in \mathbb{N}$. Without loss of generality, and by passing to a subsequence of it if necessary, we may also assume that $\big(\mu(\lambda_n)\big)_n$ is non-increasing. By (b), we have that $I(\theta,\lambda_n) \to I(\theta,\lambda)$, which clearly contradicts the second part of (4.3.4). ■

Remark 4.3.5: Under the Assumption 4.3.3, if λ, λ' are such that $\mu(\theta) < \mu(\lambda) = \mu(\lambda')$, then $I(\theta,\lambda) = I(\theta,\lambda')$. Indeed, the constant sequence $(\lambda_n)_n$ for which $\lambda_n = \lambda$ for all $n \in \mathbb{N}$, has the property that $(\mu(\lambda_n))_n$ is non-increasing with $\mu(\lambda_n) \to \mu(\lambda) = \mu(\lambda')$, so $(I(\theta,\lambda_n))_n$ converges to both $I(\theta,\lambda)$ and $I(\theta,\lambda')$, by part (b) of the previous Proposition.

Assumption 4.3.3 involved a "compatibility" condition between the Kullback-Leibler divergence and the expectation function μ. The second assumption they impose, involves a condition between the parameter space Θ and μ:

[6] We use the symbol $a_n \downarrow a$ to denote a sequence of real numbers $(a_n)_n$ which is nonincreasing and convergent to a.

Assumption 4.3.6: *The set $\Theta \subseteq \mathbb{R}$ is such that for every $\lambda \in \Theta$ and every $\delta > 0$, there exists some $\lambda' \in \Theta$ such that*

$$\mu(\lambda) < \mu(\lambda') < \mu(\lambda) + \delta. \tag{4.3.5}$$

This is a density assumption that can be equivalently reformulated as follows:

Assumption 4.3.7: *Let $A = \mu(\Theta) \subseteq \mathbb{R}$ denote the image of Θ under $\mu : \Theta \to \mathbb{R}$. Then, for every $\varepsilon > 0$ and every $a \in A$, we have that*

$$(a, a+\varepsilon) \cap A \neq \emptyset. \tag{4.3.6}$$

In particular, every point of A is an accumulation point of it. Indeed, let $a \in A$ and apply (4.3.6) for $\varepsilon_n < \frac{1}{n}$ to inductively construct a strictly decreasing sequence $(a_n)_n$ in A with $a_n \to a$. Equivalently, A is dense in itself, meaning that it contains no isolated points.[7] We therefore have the following proposition:

Proposition 4.3.8: *A set $A \subseteq \mathbb{R}$ satisfies (4.3.6) if and only if for every $a \in A$ there exists some strictly decreasing sequence $(a_n)_n$ in A with $a_n \downarrow a$.*

So, Assumption 4.3.6 can be restated again:

Assumption 4.3.9: *For every $\lambda \in \Theta$, there exists some sequence $(\lambda_n)_n$ in Θ such that $(\mu(\lambda_n))_n$ is strictly decreasing with $\mu(\lambda_n) \downarrow \mu(\lambda)$.*

At this point, we should stress out how Assumption 4.3.6 and Proposition 4.3.4 (b) complement with each other: The former asserts that it is always possible to approach any $\mu(\lambda)$ with a strictly decreasing sequence $(\mu(\lambda_n))_n$, while the latter allows us to conclude that $I(\theta, \lambda_n) \to I(\theta, \lambda)$.

During the proof of the main theorem, we will use the fact that when a sequence of random variables $(X_n)_n$ has the property that $\frac{X_n}{n} \to a > 0$ almost surely, then $\frac{\max_{k=1,\dots,n} X_k}{n}$ also converges to a almost surely; a fact that follows from the lemma below:

Lemma 4.3.10: *Let $(a_n)_n$ be a sequence of real numbers such that $\frac{a_n}{n} \to a \geq 0$, and let $(M_n)_n$ denote the sequence of its partial maxima, $M_n = \max_{k=1,\dots,n}\{a_k\}$. Then $\frac{M_n}{n} = \frac{\max_{k=1,\dots,n}\{a_k\}}{n} \to a$.*

Proof. Suppose first that $a = 0$. Let $\varepsilon > 0$ and pick some $n_0 \in \mathbb{N}$ such that $\frac{|a_n|}{n} < \frac{\varepsilon}{2}$ for every $n \geq n_0$. Then

$$\begin{aligned}
\frac{M_n}{n} &\leq \frac{M_{n_0} + \max\{|a_{n_0+1}|, \dots, |a_n|\}}{n} = \frac{M_{n_0}}{n} + \frac{\max\{|a_{n_0+1}|, \dots, |a_n|\}}{n} \\
&= \frac{M_{n_0}}{n} + \frac{|a_{k_n}|}{n} \quad \text{for some } k_n \in \{n_0+1, \dots, n\} \\
&= \frac{M_{n_0}}{n} + \frac{k_n}{n}\frac{|a_{k_n}|}{k_n} \\
&\leq \frac{M_{n_0}}{n} + \frac{\varepsilon}{2}
\end{aligned}$$

[7] However, not every dense in itself set satisfies (4.3.6); for example $A = [0, 1]$.

for $n \geq n_0$. Pick also an $n_1 \in \mathbb{N}$ such that $\frac{M_{n_0}}{n} < \frac{\varepsilon}{2}$ for every $n \geq n_1$. Then for every $n \geq \max\{n_0, n_1\}$, we have that $\frac{M_n}{n} \leq \varepsilon$, so $\limsup_n \frac{M_n}{n} \leq 0$. On the other hand, $\frac{M_n}{n} \geq \frac{a_n}{n} \to 0$, so $\frac{M_n}{n} \to 0$.

Suppose now that $a > 0$. Set $b_n = a_n - na$ and

$$M_n^b = \max_{k=1,\dots,n} \{|b_k|\} = \max_{k=1,\dots,n} \{|a_k - ka|\}.$$

Since $\frac{b_n}{n} = \frac{a_n}{n} \to a$, by the previous step we have that $\frac{M_n^b}{n} \to 0$. Notice that

$$\begin{aligned} \frac{M_n^b}{n} &= \frac{\max_{k=1,\dots,n}\{|a_k - ka|\}}{n} \\ &\geq \frac{\max_{k=1,\dots,n}\{a_k - ka\}}{n} \\ &\geq \frac{\max_{k=1,\dots,n}\{a_k - na\}}{n} \\ &= \frac{\max_{k=1,\dots,n}\{a_k\}}{n} - a, \end{aligned}$$

with the LHS tending to zero. So $\limsup_n \frac{M_n^a}{n} \leq a$. Additionally,

$$\frac{M_n^a}{n} - a \geq \frac{a_n}{n} - a \to 0,$$

from which we conclude that $\liminf_n \frac{M_n^a}{n} \geq a$. ■

Suppose that we have K bandits with parameters $\theta_1, \dots, \theta_K \in \Theta$ and expectations $\mu(\theta_1), \dots, \mu(\theta_K)$ respectively. For every $j \in \{1, \dots, K\}$, we partition the set $\Theta^k = \{(\theta_1, \dots, \theta_K) : \theta_i \in \Theta \ \forall i\}$ into $\Theta^k = \Theta_j \cup \Theta_j^* \cup \Theta_j^{**}$, where

$$\begin{aligned} \Theta_j &= \left\{\theta = (\theta_1, \dots, \theta_K) : \mu(\theta_j) < \max_{i \neq j} \mu(\theta_i)\right\}, \\ \Theta_j^* &= \left\{\theta = (\theta_1, \dots, \theta_K) : \mu(\theta_j) > \max_{i \neq j} \mu(\theta_i)\right\}, \\ \Theta_j^{**} &= \left\{\theta = (\theta_1, \dots, \theta_K) : \mu(\theta_j) = \max_{i \neq j} \mu(\theta_i)\right\}. \end{aligned}$$

The set Θ_j contains the parameters for which j is not the best arm, the set Θ_j^* the parameters for which j is the unique best arm, whereas the last one the parameters for which j is the best, but not the unique best.

Theorem 4.3.11: [LR85, Theorem 2] *Suppose that the families of distributions of the arms satisfy Assumptions 4.3.3 and 4.3.6. Fix an index $j \in \{1, \dots, K\}$ and let ϕ be any rule such that for every $\theta \in \Theta_j^*$,*

$$\sum_{i \neq j} \mathbb{E}_\theta[T_n(i)] = o(n^a) \ \text{ for every } a > 0. \tag{4.3.7}$$

Then, for every $\theta \in \Theta_j$ and every $\varepsilon > 0$,

$$\lim_n P\left[T_n(j) \geq \frac{(1-\varepsilon)\ln n}{I(\theta_j,\theta^*)}\right] = 1 \text{ and} \tag{4.3.8}$$

$$\liminf_n \mathbb{E}\left[\frac{T_n(j)}{\ln n}\right] \geq \frac{1}{I(\theta_j,\theta^*)}. \tag{4.3.9}$$

Proof. Suppose that $j = 1$, $\theta \in \Theta_1$ and $\theta^* = \theta_2$ and fix a $\delta \in (0,1)$. Since $i = 2$ is the unique best arm, we have that $\mu(\theta_2) > \mu(\theta_1)$. There exists some $\lambda \in \Theta$ such that

$$\mu(\lambda) > \mu(\theta_2) \quad \text{and} \quad |I(\theta_1,\lambda) - I(\theta_1,\theta_2)| < \delta I(\theta_1,\theta_2). \tag{4.3.10}$$

Indeed, if we set $\varepsilon = \delta I(\theta_1,\theta_2)$ and apply Proposition 4.3.3, we obtain a $\delta' > 0$ such that (4.3.3) holds, namely $|I(\theta_1,\lambda) - I(\theta_1,\theta_2)| < \delta I(\theta_1,\theta_2)$, whenever $\mu(\theta_2) \leq \mu(\lambda) \leq \mu(\theta_2) + \delta'$. By the density Assumption 4.3.6, there exists some $\lambda \in \Theta$ with $\mu(\theta_2) < \mu(\lambda) < \mu(\theta_2) + \delta'$, from which we obtain (4.3.10).

We define a new parameter vector $\gamma = (\lambda, \theta_2, \ldots, \theta_K)$. Since $\mu(\lambda) > \mu(\theta_2)$, we have that $\gamma \in \Theta_1^*$. Additionally,

$$n = \sum_{i=1}^n \mathbb{E}_\gamma[T_n(i)] = \mathbb{E}_\gamma[T_n(1)] + \sum_{i=2}^n \mathbb{E}_\gamma[T_n(i)],$$

which implies that

$$\mathbb{E}_\gamma[n - T_n(1)] = \sum_{i=2}^n \mathbb{E}_\gamma[T_n(i)] = o(n^a) \text{ for all } a \in (0,\delta) \tag{4.3.11}$$

by the asymptotic property (4.3.7), which also holds for the parameter vector γ. Using the Markov inequality, we can find a lower bound for $\mathbb{E}_\gamma[n - T_n(1)]$ with respect to the probability of the events $A_n = \left[T_n(1) < \frac{(1-\delta)\ln n}{I(\theta_1,\lambda)}\right]$:

$$\begin{aligned}\mathbb{E}_\gamma[n - T_n(1)] &= \int_{A_n} (n - T_n(1))\, dP_\gamma + \int_{A_n^c} (n - T_n(1))\, dP_\gamma \\ &\geq \int_{A_n} (n - T_n(1))\, dP_\gamma,\end{aligned}$$

as $n - T_n(1) \geq 0$. On the set A_n we have that $n - T_n(1) > n - \frac{(1-\delta)\ln n}{I(\theta_1,\lambda)}$, so for every $a \in (0,\delta)$,

$$o(n^a) = \mathbb{E}_\gamma[n - T_n(1)] \geq n - \frac{(1-\delta)\ln n}{I(\theta_1,\lambda)} P_\gamma(A_n). \tag{4.3.12}$$

Let $Y_1, Y_2, \ldots$ denote successive observations from the first arm and set

$$L_m = \sum_{i=1}^m \ln\left(f(Y_i;\theta_1)/f(Y_i,\lambda)\right). \tag{4.3.13}$$

Let also

$$C_n = A_n \cap \left[L_{T_n(1)} \leq (1-a)\ln n\right]. \tag{4.3.14}$$

For the sets A_n we have that $\frac{P_\gamma(A_n)}{n^{a-1}} - \frac{\ln n P_\gamma(A_n)}{n^a}\frac{1-\delta}{I(\theta_1,\lambda)} \to 0$, and since the second term of this sequences converges to zero, we conclude that $P_\gamma(C_n) \le P_\gamma(A_n) = o(n^{a-1})$.

For every k-tuple $\tilde{n} = (n_1, \dots, n_K)$, let

$$A_{\tilde{n}} = \left[T_n(1) = n_1, \dots, T_n(K) = n_K \text{ and } L_{n_1} \le (1-a)\ln n\right].$$

Then

$$\begin{aligned} P_\gamma(A_{\tilde{n}}) &= \int_{A_{\tilde{n}}} \prod_{i=1}^{n_1} f(y_i;\lambda) \prod_{i=1}^{n_2} f(y_i^2;\theta_2)\cdots\prod_{i=1}^{n_K} f(y_i^K;\theta_K)d\tilde{y} \\ &= \int_{A_{\tilde{n}}} \frac{\prod_{i=1}^{n_1} f(y_i;\lambda)}{\prod_{i=1}^{n_1} f(y_i;\theta_1)} \prod_{i=1}^{n_1} f(y_i;\theta_1) \prod_{i=1}^{n_2} f(y_i^2;\theta_2)\cdots\prod_{i=1}^{n_K} f(y_i^K;\theta_K)d\tilde{y} \\ &= \int_{A_{\tilde{n}}} \frac{\prod_{i=1}^{n_1} f(y_i;\lambda)}{\prod_{i=1}^{n_1} f(y_i;\theta_1)} dP_\theta. \end{aligned} \tag{4.3.15}$$

The inequality $L_{n_1} \le (1-a)\ln n$ implies that $\sum_{i=1}^{n_1} \ln\frac{f(y_i;\theta_1)}{f(y_i;\lambda)} \le (1-a)\ln n$ and thus, $\ln\prod_{i=1}^{n_1}\frac{f(y_i;\lambda)}{f(y_i;\theta_1)} \le -(1-a)\ln n$, so (4.3.15) yields that

$$P_\gamma(A_{\tilde{n}}) = \int_{A_{\tilde{n}}} \frac{\prod_{i=1}^{n_1} f(y_i;\lambda)}{\prod_{i=1}^{n_1} f(y_i;\theta_1)} dP_\theta \ge P_\theta(A_{\tilde{n}}) e^{-(1-a)\ln n} = n^{a-1} P_\theta(A_{\tilde{n}})$$

for every $A_{\tilde{n}}$. Since C_n can be written as the disjoint union of the sets $A_{\tilde{n}}$ for $\tilde{n}$ such that $n_1 + \ldots + n_K = n$, the same inequality holds for the sets C_n as well:

$$P_\theta(C_n) \le n^{1-a} P_\gamma(C_n) \to 0 \quad \text{for all } n. \tag{4.3.16}$$

By the Strong Law of Large Numbers, $\frac{L_m}{m} \to I(\theta_1,\lambda) > 0$ almost surely with respect to P_θ, and by Lemma 4.3.10, p. 92, so does $\frac{\max_{i=1,\dots,m} L_i}{m}$.

For every n, we set $b_n = \frac{(1-\delta)\ln n}{I(\theta_1,\lambda)}$. We have that

$$\begin{aligned} P_\theta\left[\max_{i=1,\dots,[b_n]} L_i > (1-a)\ln n\right] &= P_\theta\left[\max_{i=1,\dots,[b_n]} \frac{L_i}{[b_n]} > \frac{(1-a)\ln n}{[b_n]}\right] \\ &\le P_\theta\left[\max_{i=1,\dots,[b_n]} \frac{L_i}{[b_n]} > \frac{(1-a)\ln n}{(1-\delta)\ln n / I(\theta_1,\lambda)}\right] \\ &= P_\theta\left[\max_{i=1,\dots,[b_n]} \frac{L_i}{[b_n]} > (1+M)I(\theta_1,\lambda)\right] \to 0 \end{aligned}$$

almost surely in P_θ since $\max_{i=1,\dots,[b_n]} \frac{L_i}{[b_n]} \to I(\theta_1,\lambda)$ a.s. and $(1+M)I(\theta_1,\lambda) > I(\theta_1,\lambda)$. Therefore, if we set $B_n = \left[L_{T_n(1)} \le (1-a)\ln n\right]$, we have that $P_\theta(B_n) \to 1$. In turn, this implies that $\lim_n P_\theta(A_n) = \lim_n P_\theta(A_n \cap B_n) = \lim P_\theta(C_n) = 0$.

By (4.3.10), $\frac{1}{I(\theta_1,\lambda)} \ge \frac{1}{(1+\delta)I(\theta_1,\theta_2)}$, so

$$P_\theta\left[T_n(1) < \frac{(1-\delta)\ln n}{(1+\delta)I(\theta_1,\theta_2)}\right] \le P_\theta\left[T_n(1) < \frac{(1-\delta)\ln n}{I(\theta,\lambda)}\right] \to 0 \tag{4.3.17}$$

and (4.3.8) is established.

Lastly, we show that $\liminf_n \mathbb{E}\left[\frac{T_n(j)}{\ln n}\right] \geq \frac{1}{I(\theta_j,\theta^*)}$. Let $\varepsilon > 0$. By the Markov inequality,

$$\begin{aligned}\mathbb{E}\left[\frac{T_n(j)}{\ln n}\right] &= \int_{\left[\frac{T_n}{\ln n} \geq \frac{1-\varepsilon}{I(\theta_j,\theta^*)}\right]} \frac{T_n(j)}{\ln n} dP + \int_{\left[\frac{T_n}{\ln n} < \frac{1-\varepsilon}{I(\theta_j,\theta^*)}\right]} \frac{T_n(j)}{\ln n} dP \\ &\geq \frac{1-\varepsilon}{I(\theta_j,\theta^*)} P\left[\frac{T_n}{\ln n} \geq \frac{1-\varepsilon}{I(\theta_j,\theta^*)}\right],\end{aligned}$$

so, by taking limits we obtain that

$$\liminf_n \mathbb{E}\left[\frac{T_n(j)}{\ln n}\right] \geq \liminf_n \frac{1-\varepsilon}{I(\theta_j,\theta^*)} P\left[\frac{T_n}{\ln n} \geq \frac{1-\varepsilon}{I(\theta_j,\theta^*)}\right] = \frac{1-\varepsilon}{I(\theta_j,\theta^*)}. \quad (4.3.18)$$

Since (4.3.18) holds for every $\varepsilon > 0$, by letting $\varepsilon \to 0$, we have that

$$\liminf_n \mathbb{E}\left[\frac{T_n(j)}{\ln n}\right] \geq \frac{1}{I(\theta_j,\theta^*)},$$

as we wanted. ■

Theorem 4.3.12: *Suppose that the families of distributions of the arms satisfy Assumptions 4.3.3 and 4.3.6. Let ϕ be a rule whose regret satisfies the property that*

$$R_n(\theta) = o(n^a) \quad (4.3.19)$$

for every $\theta = (\theta_1, \ldots, \theta_K) \in \Theta^K$ and $a > 0$. Then for every θ such that the $\mu(\theta_i)$ are not all equal, we have that

$$\liminf_n \frac{R_n(\theta)}{\ln n} \geq \sum_{i:\mu(\theta_i)<\mu^*} \frac{\mu^* - \mu(\theta_i)}{I(\theta_i,\theta^*)}. \quad (4.3.20)$$

Proof. We will use the elementary property from real analysis, according to which for any two real sequences $(a_n)_n$, $(b_n)_n$,

$$\liminf_n (a_n + b_n) \geq \liminf_n a_n + \liminf_n b_n.$$

By applying it for R_n, we obtain that

$$\begin{aligned}\liminf_n \frac{R_n(\theta)}{\ln n} &= \liminf_n \frac{\sum_{i=1}^K \left(\mu^* - \mu(\theta_i)\right) \mathbb{E}_\theta[T_n(i)]}{\ln n} \\ &\geq \sum_{i:\mu(i)<\mu^*} (\mu^* - \mu(\theta_i)) \liminf_n \frac{\mathbb{E}_\theta[T_n(i)]}{\ln n} \\ &\geq \sum_{i:\mu(i)<\mu^*} \frac{\mu^* - \mu(\theta_i)}{I(\theta_i,\theta^*)}\end{aligned}$$

which establishes 4.3.20. ■

4.3.3 *An algorithm which achieves the theoretical lower bound*

In the same article, Lai and Robbins proposed an algorithm which achieves the optimal lower bound. The principle behind their idea is called *optimism in the face of uncertainty* and can be summarized as follows: Recall that in the classical Robbins Rule, at each stage we compared the populations sample means and chose the one with the largest. In this algorithm instead, the arms which have not been explored enough are not represented by their sample mean, but by an upper confidence bound of it. By comparing the sample means of some arms with the upper confidence bounds of some others, we are giving the algorithm the incentive to explore more.

This principle of regarding the upper confidence bounds to be better indicator of what the expectation could be, than the sample mean, will be a common theme in all of the algorithms we present in this chapter. However, it is important to realize that the difficult task of finding a balance between exploitation and exploration, has not been addressed yet, but has only been reformulated into: *How large should the confidence intervals be?*

The larger the intervals, the more our algorithm tends to explore, but the more it explores, the less it exploits. Optimism, as a general principle, is a first step towards the right direction, but when actually employing it, one needs to also be prepared for the difficult task of quantifying exactly *how optimistic* he is willing to be. As one can already imagine, determining the upper confidence bounds requires extremely delicate work.

Lai and Robbins tackle it by listing certain axioms the upper confidence functions g_{ni} need to satisfy. They also allow for the sample mean to be replaced by some more general statistic h satisfying another list of axioms, but we should keep in mind that both of these classes of functions are intended to generalize the notions of upper confidence bounds and sample means respectively.

Assumption 4.3.13: *For every $n \in \mathbb{N}$ and $i = 1, \dots, n$, there exist Borel functions $g_{ni} : \mathbb{R}^i \to \mathbb{R}$, called the upper confidence bounds, such that for every $\theta \in \Theta$, every $r < \mu(\theta)$ and every λ with $\mu(\lambda) > \mu(\theta)$,*

$$\lim_{\varepsilon \downarrow 0} \left(\limsup_n \sum_{i=1}^{n} \frac{P_\theta \left[g_{ni}(Y_1, \dots, Y_i) \geq \mu(\lambda) - \varepsilon \right]}{\ln n} \right) \leq \frac{1}{I(\theta, \lambda)}, \tag{4.3.21}$$

$$P_\theta \left[r \leq g_{ni}(Y_1, \dots, Y_i) \text{ for all } i \leq n \right] = 1 - o(n^{-1}) \text{ and} \tag{4.3.22}$$

$$g_{ni} \text{ is nondecreasing with respect to } n \geq i \text{ for every fixed } i. \tag{4.3.23}$$

The existence of such functions is not straightforward and [LR85, Section 4] contains some general rules of constructing them, as well as specific examples for several known distributions. However, things are a little easier when constructing the expectation estimates:

Assumption 4.3.14: *There exist functions $h_i : \mathbb{R}^i \to \mathbb{R}$, called point estimates of the expectation, which satisfy the following properties:*

$$h_i \leq g_{ni} \text{ for all } n \geq i \tag{4.3.24}$$

and

$$P_\theta\left[\max_{\delta n\le i\le n}|h_i(Y_1,\dots,Y_i)-\mu(\theta)|>\varepsilon\right]=o(n^{-1}) \tag{4.3.25}$$

for all $\theta\in\Theta,\varepsilon>0$ *and* $\delta\in(0,1)$.

An example of a statistic that satisfies (4.3.25) is the sample mean $h(Y_1,\dots,Y_n)=\frac{Y_1+\dots+Y_n}{n}=\frac{S_n}{n}$. We can show this, using the Kolmogorov and Marcinkiewicz - Zygmund inequalities.

Theorem 4.3.15 (Kolmogorov's Maximal Inequality [Gut13, Thm 1.6, p. 122]**):** *Let* $(X_n)_n$ *be a sequence of independent random variables with* $\mathbb{E}[X_n]=0$ *and* $V(X_n)<\infty$ *for all* $n\in\mathbb{N}$. *Then, for every* $\varepsilon>0$,

$$P\left[\max_{1\le k\le n}|S_k|>\varepsilon\right]\le\frac{\sum_{k=1}^n V(X_k)}{\varepsilon^2}=\frac{V(S_n)}{\varepsilon^2}=\frac{\mathbb{E}[S_n^2]}{\varepsilon^2}. \tag{4.3.26}$$

If, in addition, the sequence $(X_n)_n$ *is identically distributed, then*

$$P\left[\max_{1\le k\le n}|S_k|>\varepsilon\right]\le\frac{nV(X_1)}{\varepsilon^2}=\frac{n\mathbb{E}[X_1^2]}{\varepsilon^2}. \tag{4.3.27}$$

We actually need a strengthened version of Kolmogorov's Inequality (for a proof, see [Saa17] or [Doo90, p. 317]), where the exponent 2 is replaced by any $p\ge1$:

Theorem 4.3.16 (Doob-Kolmogorov): *Let* $(X_n)_n$ *be a sequence of independent random variables with* $\mathbb{E}[X_n]=0$ *and* $V(X_n)<\infty$ *for all* $n\in\mathbb{N}$. *Then, for every* $\varepsilon>0$ *and* $p\ge1$,

$$P\left[\max_{1\le k\le n}|S_k|>\varepsilon\right]\le\frac{\mathbb{E}[|S_n|^p]}{\varepsilon^p}. \tag{4.3.28}$$

Theorem 4.3.17 (Marcinkiewicz-Zygmund Inequality [MZ37]**):** *Let* $(X_n)_n$ *be a sequence of independent random variables with* $\mathbb{E}[X_n]=0$ *and* $\mathbb{E}|X_n|^p<\infty$ *for all* $n\in\mathbb{N}$ *for some* $p\ge1$. *Then, there exist constants* A_p,B_p *depending only on* p, *such that*

$$A_p\mathbb{E}\left[\left(\sum_{k=1}^n X_k^2\right)^{p/2}\right]\le\mathbb{E}|S_n|^p\le B_p\mathbb{E}\left[\left(\sum_{k=1}^n X_k^2\right)^{p/2}\right] \tag{4.3.29}$$

for all n. *If, in addition, the sequence* $(X_n)_n$ *is identically distributed, then there exists a constant* B_p^* *depending only on* p, *such that*

$$\mathbb{E}|S_n|^p\le\begin{cases}nB_p^*\mathbb{E}|X_1|^p, & 1\le p\le2,\\ n^{p/2}B_p\mathbb{E}|X_1|^{p/2}, & p\ge2.\end{cases} \tag{4.3.30}$$

Proposition 4.3.18: [CR75, p. 55] *Suppose that* $(Y_n)_n$ *is an i.i.d. sequence for which there exists some* $p>2$ *such that* $\mathbb{E}[|Y_1|^p]<\infty$. *Let also* μ *denote their common expectation. Then the sample means* $\overline{Y}_n=\frac{Y_1+\dots+Y_n}{n}$ *satisfy property (4.3.25), namely*

$$P_\theta\left[\max_{\delta n\le i\le n}|\overline{Y}_i-\mu|>\varepsilon\right]=o(n^{-1})$$

for all $\varepsilon>0$ *and* $\delta\in(0,1)$.

Proof. For the sequence $X_n = Y_n - \mu$, we have that that $\mathbb{E}[X_n] = 0$ for all n and $\mathbb{E}[|X_1|^p] < \infty$. Additionally, $\frac{S_n}{n} := \frac{X_1 + \ldots + X_n}{n} = \overline{Y}_n - \mu$. A trivial, but extremely important observation, is that

$$\left[\max_{\delta n \leq i \leq n} \left|\frac{S_i}{i}\right| > \varepsilon\right] \subseteq \left[\max_{\delta n \leq i \leq n} \left|\frac{S_i}{\delta n}\right| > \varepsilon\right].$$

This allows the denominators appearing in these events to be absorbed by ε, so that we can apply Kolmogorov's Inequality afterwards.

$$\begin{aligned}
P_\theta\left[\max_{\delta n \leq i \leq n} |\overline{Y}_i - \mu| > \varepsilon\right] &= P\left[\max_{\delta n \leq i \leq n} \left|\frac{S_i}{i}\right| > \varepsilon\right] \\
&\leq P\left[\max_{\delta n \leq i \leq n} \left|\frac{S_i}{\delta n}\right| > \varepsilon\right] \\
&= P_\theta\left[\max_{1 \leq i \leq n} |S_i| > n\varepsilon\delta\right] \\
&\leq \frac{\mathbb{E}[S_n^p]}{n^p \varepsilon^p \delta^p} && \text{Kolmogorov's Inequality} \\
&\leq \frac{n^{p/2} B_p \mathbb{E}|X_1|^{p/2}}{n^p \varepsilon^p \delta^p} && \text{Marcinkiewicz-Zygmund} \\
&= \frac{C}{n^{p/2}}.
\end{aligned}$$

By multiplying with n, we obtain that

$$nP_\theta\left[\max_{\delta n \leq i \leq n} |\overline{Y}_i - \mu| > \varepsilon\right] \leq \frac{nC}{n^{p/2}} \to 0,$$

since $\frac{p}{2} > 1$.

■

To return to the multi-armed bandit setting, suppose that we have K-bandits with densities $f(x;\theta_1), \ldots, f(x;\theta_K)$ and let ϕ be a rule of sampling from them. For each $j \in \{1, \ldots, K\}$, let $T_n(j)$ denote the number of times ϕ sampled from the j-bandit during the first n-stages, namely

$$T_n(j) = \#\left\{i \in \{1, \ldots, n\} : \phi(i) = j\right\}. \tag{4.3.31}$$

We denote the successive observations from the j-bandit during the first n-stages as $Y_{j1}, \ldots, Y_{jT_n(j)}$, and in accordance with our previous discussion, we define the mean estimates and upper confidence bounds of $\mu(\theta_j)$ based on this sample as

$$\tilde{\mu}_n(j) = h_{T_n(j)}\left(Y_{j1}, \ldots, Y_{jT_n(j)}\right) \text{ and} \tag{4.3.32}$$

$$U_n(j) = g_{nT_n(j)}\left(Y_{j1}, \ldots Y_{jT_n(j)}\right) \tag{4.3.33}$$

respectively.

Lai and Robbins use the point and upper confidence bound estimates to construct their algorithm: We fix a $\delta \in (0, 1/k)$ and define their rule ϕ adaptively

as follows: During the first K-rounds, we pick each arm exactly once. For every $n \geq K$, we set

$$I_n = \left\{ j \in \{1, \dots, K\} : T_n(j) \geq \delta n \right\}.$$

By the Pigeonhole Principle, these sets are always nonempty. Indeed, if some I_n was empty, then $T_n(j) < n\delta$ should hold for every $j = 1, \dots, K$. But then, $n = \sum_{j=1}^{n} T_n(j) < nk\delta < n$, a contradiction. For all such n's, we also set

$$j_n = \arg\max \left\{ \tilde{\mu}_n(j) : j \in I_n \right\} \quad \text{and} \quad \tilde{\mu}_n(j_n) = \max \left\{ \tilde{\mu}_n(j) : j \in I_n \right\}.$$

Suppose that the rule has just drawn its n-th value with $n \geq K$ and let $j \in \{1, \dots, K\}$ be such that $n + 1 \equiv j \mod K$. Then

$$\phi(n+1) = \begin{cases} j, & \text{if } \tilde{\mu}_n(j_n) \leq U_n(j), \\ j_n, & \text{otherwise.} \end{cases} \tag{4.3.34}$$

So, the algorithm compares the mean estimate of the current leader among the bandits which have been played enough times ($\geq \delta n$), with the upper confidence bound of the newly proposed bandit, and chooses accordingly.

The Lai-Robbins Algorithm	
Step 1	For $k = 1, \dots, K$ set $\phi(k) = k$.
Step 2	Suppose that $\phi(n)$ has been drawn for $n \geq K$ and let j be such that $n + 1 \equiv j \mod K$. Set $I_n = \{k \in \{1, \dots, K\} : T_n(k) \geq \delta n\}$, $j_n = \arg\max \{\tilde{\mu}_n(k) : k \in I_n\}$ and $\tilde{\mu}_n(j_n) = \max \{\tilde{\mu}_n(k) : k \in I_n\}$.
Step 3	If $\tilde{\mu}_n(j_n) \leq U_n(j)$, set $\phi(n+1) = j$, otherwise set $\phi(n+1) = j_n$.
Step 4	Go to Step 2.

Table 6: The Lai-Robbins Algorithm [LR85].

We mention two lemmas which will be used in the proof of the convergence rate.

Lemma 4.3.19: *If for the real sequence $(a_n)_n$ we have that $\limsup_n a_n \leq M$, then $a_n \leq M + o(1)$.*

Proof. Clearly $a_n \leq M + a_n - M$ for all $n \in \mathbb{N}$. Set

$$b_n = \begin{cases} a_n - M, & \text{when } a_n \geq M, \\ 0, & \text{when } a_n < M. \end{cases}$$

It is easy to confirm that $a_n \leq M + b_n$ for all n and that $b_n \to 0$. ∎

Lemma 4.3.20: *If $(a_n)_n$ is $o(n^{-1})$ then $\left(\sum_{n=1}^{N} a_n\right)_N = o(\ln N)$.*

Proof. We will show that for every $\varepsilon > 0$ there exists some $N_0 \in \mathbb{N}$ such that $\sum_{n=1}^{N} a_n \le \varepsilon \ln N$ for every $N \ge N_0$. Since $(a_n)_n$ is $o(n^{-1})$, there exists some n_0 such that $a_n \le \frac{\varepsilon}{2n}$ for every $n \ge n_0$. By the elementary inequality $\sum_{n=1}^{N} \frac{1}{n} \le 1 + \int_1^N \frac{1}{x} dx = 1 + \ln N$, we obtain that

$$\sum_{n=1}^{N} a_n \le \sum_{n=1}^{n_0-1} a_n + \frac{\varepsilon}{2} \sum_{n=n_0}^{N} \frac{1}{n} \le \sum_{n=1}^{n_0-1} a_n + \frac{\varepsilon}{2} (\ln N - \ln(n_0 - 1)) \le \sum_{n=1}^{n_0-1} a_n + \frac{\varepsilon}{2} \ln N$$

for every $N \ge n_0$. Pick some n_1 such that $\sum_{n=1}^{n_0-1} a_n \le \frac{\varepsilon}{2} \ln n_1$. Then for every $N \ge \max\{n_0, n_1\}$, the inequality $\sum_{n=1}^{N} a_n \le \varepsilon$ holds. ∎

Theorem 4.3.21: [LR85, Theorem 3] *Suppose that $I(\theta, \lambda)$ satisfies Assumption 4.3.3 and let ϕ be the Lai-Robbins rule. Then for every $\theta = (\theta_1, \dots, \theta_K)$ and every j such that $\mu(\theta_j) < \mu(\theta^*)$,*

$$\mathbb{E}_\theta[T_n(j)] \le \left(\frac{1}{I(\theta_j, \theta^*)} + o(1) \right) \ln n. \tag{4.3.35}$$

If, in addition, Θ satisfies the density Assumption 4.3.6, then $\mathbb{E}_\theta[T_n(j)] \sim \frac{\ln n}{I(\theta_j, \theta^*)}$ for every such j, and the regret R_n of ϕ satisfies

$$R_n \sim \left(\sum_{j : \mu(\theta_j) < \mu^*} \frac{\mu^* - \mu(\theta_j)}{I(\theta_j, \theta^*)} \right) \ln n \tag{4.3.36}$$

Proof. Let L denote the set of optimal arms, $L = \{l \in \{1, \dots, L\} : \mu(\theta_l) = \mu(\theta^*)\}$, and set $\varepsilon < \frac{1}{2} \min\{\mu(\theta^*) - \mu(\theta_j) : j \notin L\}$. For every $j \notin L$ and $N \in \mathbb{N}$, we can partition the event $T_N(j)$ as follows:

$$\begin{aligned}
T_N(j) &= \#\{n \in \{1, \dots, N\} : \phi(n) = j\} \\
&= 1 + \#\{n \in \{1, \dots, N-1\} : j_n \in L, |\tilde{\mu}_n(j_n) - \mu(\theta^*)| \le \varepsilon, \phi(n+1) = j\} + \\
&\quad + \#\{n \in \{1, \dots, N-1\} : j_n \in L, |\tilde{\mu}_n(j_n) - \mu(\theta^*)| > \varepsilon, \phi(n+1) = j\} + \\
&\quad + \#\{n \in \{1, \dots, N-1\} : j_n \notin L, \phi(n+1) = j\} \\
&\le 1 + \underbrace{\#\{n \in \{1, \dots, N-1\} : j_n \in L, |\tilde{\mu}_n(j_n) - \mu(\theta^*)| \le \varepsilon, \phi(n+1) = j\}}_{A_1} + \\
&\quad + \underbrace{\#\{n \in \{1, \dots, N-1\} : j_n \in L, |\tilde{\mu}_n(j_n) - \mu(\theta^*)| > \varepsilon\}}_{A_2} + \\
&\quad + \underbrace{\#\{n \in \{1, \dots, N-1\} : j_n \notin L\}}_{A_3}.
\end{aligned}$$

In order to bound A_1, notice that since $|\tilde{\mu}_n(j_n) - \mu(\theta^*)| \le \varepsilon$ and $\phi(n+1) = j$, then it has to be that $g_{ni}(Y_{j1}, \dots, Y_{ji}) > \mu(\theta^*) - \varepsilon$, where i is the number of observation taken from the j-bandit after the first $n \le N-1$ rounds. Furthermore, g_{ni} is increasing with respect to n, so

$$A_1 \le 2 + \#\{1 \le i \le N-1 : g_{Ni}(Y_{j1}, \dots, Y_{ji}) \ge \mu(\theta^*) - \varepsilon\}. \tag{4.3.37}$$

Let $A_i^\varepsilon = \left[g_{ni}(Y_{j1},\dots,Y_{ji}) \geq \mu(\lambda) - \varepsilon\right]$. By (4.3.21), we have that

$$\lim_{\varepsilon \downarrow 0}\left(\limsup_n \sum_{i=1}^n \frac{P_\theta\left(A_i^\varepsilon\right)}{\ln n}\right) \leq \frac{1}{I(\theta,\lambda)},$$

so for every $\rho > 0$, there exists some $\varepsilon_0 > 0$ such that

$$\limsup_n \sum_{i=1}^n \frac{P_\theta\left(A_i^\varepsilon\right)}{\ln n} \leq \frac{1}{I(\theta_j,\theta^*)} + \frac{\rho}{I(\theta_j,\theta^*)}$$

for every $0 < \varepsilon \leq \varepsilon_0$. By Lemma 4.3.19,

$$\sum_{i=1}^n \frac{P_\theta\left(A_i^\varepsilon\right)}{\ln n} \leq \frac{1+\rho}{I(\theta_j,\theta^*)} + o(1),$$

which implies that

$$\sum_{i=1}^n P_\theta\left(A_i^\varepsilon\right) \leq \frac{1+\rho+o(1)}{I(\theta_j,\theta^*)} \ln n$$

for all $0 < \varepsilon < \varepsilon_0$. Lastly,

$$\mathbb{E}\left[\#\left\{1 \leq i \leq n : A_i^\varepsilon \text{ occurs}\right\}\right] = \mathbb{E}\left[\sum_{i=1}^n I_{A_i^\varepsilon}\right] = \sum_{i=1}^n P(A_i^\varepsilon) \leq \frac{1+\rho+o(1)}{I(\theta_j,\theta^*)} \ln n$$

for ε small enough.

For the A_2 term, let $B_n = \left[j_n \in L, |\tilde{\mu}_n(j_n) - \mu(\theta^*)| > \varepsilon\right]$. By (4.3.25),

$$P_\theta[B_n] \leq P_\theta\left[\max_{l \in L} \max_{\delta n \leq i \leq n} |h_i(Y_{i1},\dots,Y_{li}) - \mu(\theta^*)| > \varepsilon\right] = o(n^{-1}),$$

so, by Lemma 4.3.20,

$$\mathbb{E}_\theta[A_2] = \sum_{n=1}^{N-1} o(n^{-1}) = o(\ln N). \tag{4.3.38}$$

We will show in Lemma 4.3.22 that $\mathbb{E}_\theta[A_3]$ is also $o(\ln N)$, so combining all the above we obtain that

$$\mathbb{E}_\theta[T_n(j)] \leq \frac{1+\rho+o(1)}{I(\theta_j,\theta^*)} \ln n \quad \text{for all } \rho > 0.$$

Therefore, with an argument similar to the one in the proof of Lemma 4.3.19,

$$\mathbb{E}_\theta[T_n(j)] \leq \left(\frac{1}{I(\theta_j,\theta^*)} + o(1)\right) \ln n. \tag{4.3.39}$$

The additional part follows from (4.3.39) and (4.3.9). ■

Lemma 4.3.22: *Under the notation of the previous Theorem and its proof, let $c \in \mathbb{N}$ and for $r = 0, 1 \ldots$, define*

$$A_r = \bigcap_{j=1}^{K} \left[\max_{\delta c^{r-1} \le n \le c^{r+1}} \left| h_n(Y_{j1}, \ldots, Y_{jn}) - \mu(\theta_j) \right| \le \varepsilon \right] \text{ and}$$
$$B_r = \bigcap_{l \in L} \left[g_{ni}(Y_{l1}, \ldots, Y_{li}) \ge \mu(\theta^*) - \varepsilon, \ \forall 1 \le i \le \delta n, \ \forall c^{r-1} \le n \le c^{r+1} \right].$$

Then $P_\theta(A_r^c) = o(c^{-r})$ and $P_\theta(B_r^c) = o(c^{-r})$. Additionally, if $c > \frac{1}{1-k\delta}$, then $j_n \in L$ for all $n \in [c^r, c^{r+1}]$ on $A_r \cap B_r$, for r sufficiently large. Consequently,

$$\mathbb{E}_\theta \left[\#\{1 \le n \le N : j_n \notin L\} \right] = \sum_{n=1}^{N} P_\theta[j_n \notin L] = o(\ln N).$$

Proof. For fixed c and r as in the statement of the Lemma, we consider the sequence of intervals

$$\begin{aligned} I_1 &= [\delta c^{r-1}, c^{r-1}], \\ I_2 &= [c^{r-1}, \delta^{-1} c^{r-1}], \\ I_3 &= [\delta^{-1} c^{r-1}, \delta^{-2} c^{r-1}], \\ &\vdots \\ I_{n+1} &= [\delta^{-n+1} c^{r-1}, \delta^{-n} c^{r-1}], \end{aligned}$$

until $\delta^{-n} c^{r-1} \ge c^{r+1}$ for the first time, namely for $n = n_0 := \left[-\frac{2 \ln c}{\ln \delta} \right] + 1$. Then

$$\begin{aligned} A_r^c &= \bigcup_{j=1}^{K} \left[\max_{\delta c^{r-1} \le n \le c^{r+1}} \left| h_n(Y_{j1}, \ldots, Y_{jn}) - \mu(\theta_j) \right| > \varepsilon \right] \\ &\subseteq \bigcup_{j=1}^{K} \left(\bigcup_{l=1}^{n_0} \left[\max_{n \in I_l} \left| h_n(Y_{j1}, \ldots, Y_{jn}) - \mu(\theta_j) \right| > \varepsilon \right] \right), \quad \text{so} \\ P(A_r^c) &\le \sum_{j=1}^{K} \sum_{l=1}^{n_0} P \left[\max_{n \in I_l} \left| h_n(Y_{j1}, \ldots, Y_{jn}) - \mu(\theta_j) \right| > \varepsilon \right] \\ &= \sum_{j=1}^{K} \sum_{l=1}^{n_0} o\left(\frac{\delta^{l-1}}{c^{r-1}} \right), \quad \text{by Property 4.3.25,} \\ &= o(c^{-r}), \end{aligned}$$

since K is constant and n_0 is independent of r. Regarding B_r^c, for $t = 0, \ldots, n_0$, let $n_t = \left[\frac{c^{r-1}}{\delta^t} \right]$ and

$$D_t = \bigcap_{l \in L} \left[g_{n_t i}(Y_{l1}, \ldots, Y_{li}) \ge \mu(\theta^*) - \varepsilon, \ \forall i \le n_t \right].$$

By (4.3.22),

$$\begin{aligned} P_\theta(D_t^c) &\le \sum_{l \in L} P_\theta \left(\left[g_{n_t i}(Y_{l1}, \ldots, Y_{li}) \ge \mu(\theta^*) - \varepsilon, \ \forall i \le n_t \right]^c \right) \\ &\le \#L \cdot o(n_t^{-1}) \end{aligned}$$

$$= o(n_t^{-1})$$
$$= o(c^{-r}),$$

the last one following from the inequality $c^{r-1} - 1 \leq n_t \leq \delta^{-n_0} c^{r-1} + 1$ for $t = 0, \ldots, n_0$. Given $n \in [c^{r-1}, c^{r+1}]$ and $i \in [1, \delta n]$, there exists a $t \in \{0, \ldots, n_0 - 1\}$ such that $n_{t+1} > n \geq n_t \geq i$. Indeed, the inequality $n_{n_0} \geq c^{r+1} > c^{r-1} \geq n_0 = [c^{r-1}]$ yields the existence of some $t \in \{0, \ldots, n_0 - 1\}$ with $n_{t+1} > n \geq n_t$. By the monotonicity of the family g, we have that $g_{ni}(Y_{l1}, \ldots, Y_{li}) \geq g_{n_t i}(Y_{l1}, \ldots, Y_{li}) \geq \mu(\theta^*) - \varepsilon$ for all $l \in L$ on $\cap_{t=0}^{n_0} D_t$. Since $B_r^c \subseteq \cup_{t=0}^{n_0} D_t$ we obtain that

$$P(B_r^c) \leq \sum_{t=0}^{n_0} P(D_t) \leq (n_0 + 1) o(c^{-r}) = o(c^{-r}).$$

For the additional part, suppose that $c > \frac{1}{1-k\delta}$. Let $\nu_L(n)$ denote the number of times ϕ samples from L during the first n rounds. Then

$$\nu_L(n) = \sum_{l \in L} T_n(l) \leq \sum_{l \in L} \max_{k \in L} T_n(k) = \#L \cdot \max_{k \in L} T_n(k), \text{ so}$$
$$\frac{\nu_L(n)}{\#L} \leq \max_{k \in L} T_n(k). \tag{4.3.40}$$

We consider the round $n + 1 \equiv l \mod K$ with $l \in L$ and $c^{r-1} \leq n < c^{r+1}$. We will show that ϕ must sample from L on the event $A_r \cap B_r$. According to the algorithm, if $j_n \in L$, then the only two candidate arms to sample from during the $n+1$ round, are j_n and l, both of which belong to L, so ϕ draws from L trivially. So only the case where $j_n \notin L$ needs to be addressed.

Suppose that $j_n \notin L$. Since $T_n(j_n) \geq \delta n$ and $2\varepsilon < \mu(\theta^*) - \max_{j \notin L} \mu(\theta_j)$ (see p. 101), we have that

$$\tilde{\mu}_n(j_n) \leq \max_{j \notin L} \mu(\theta_j) + \varepsilon < \mu(\theta^*) - \varepsilon \tag{4.3.41}$$

on A_r. For the l-arm there are two possibilities, either $T_n(l) \geq \delta n$, or $T_n(l) < \delta n$. When the former holds, on A_r we have that

$$\mu(\theta^*) - \varepsilon \leq h_{T_n(l)}(Y_{l1}, \ldots, Y_{lT_n(l)}) \leq g_{nT_n(l)}(Y_{l1}, \ldots, Y_{lT_n(l)}), \tag{4.3.42}$$

so $\tilde{\mu}_n(j_n) < g_{nT_n(l)}(Y_{l1}, \ldots, Y_{lT_n(l)})$ and ϕ samples from $l \in L$ at round $n + 1$.

When $T_n(l) < \delta n$, the same is true, since

$$\mu(\theta^*) - \varepsilon \leq g_{nT_n(j)}(Y_{l1}, \ldots, Y_{lT_n(j)})$$

on the event B_r. Therefore, ϕ always samples from L on the event $A_r \cap B_r$, at every stage $n + 1$ for which $n + 1 \equiv l \mod K$ with $l \in L$ and $n \in [c^{r-1}, c^{r+1}]$.

In order to find a lower bound for $\nu_L(n)$ on $A_r \cap B_r$, we diverge slightly from the Lai-Robbins article, and additionally assume that $c \geq 2$, δ and k are chosen such that they also satisfy the relations $\delta < \frac{1}{100K}$ and $\frac{1-c^{-1}}{K} > \delta c$.[8] Observe that

[8] One can easily verify that there always exists such c, δ and k.

on the interval $[c^{r-1}, n]$, the algorithm always draws from L during the round m when $m+1 \equiv l \mod K$ with $l \in L$, something which occurs at least $\#L$ times for each cycle. Then,

$$\begin{aligned}
\nu_L(n) &\geq \#\left\{\begin{matrix}\text{Observations from } L \\ \text{on the interval } [1,K]\end{matrix}\right\} + \#\left\{\begin{matrix}\text{Observations from } L \\ \text{on the interval } [c^{r-1},n]\end{matrix}\right\} \\
&\geq \#L + \#L \cdot \#\left\{\begin{matrix}\text{Disjoint intervals of} \\ \text{length } K \text{ in } [c^{r-1},n]\end{matrix}\right\} \\
&\geq \#L + \#L \cdot \left\lfloor \frac{n-c^{r-1}}{K} \right\rfloor \\
&\geq \#L + \#L \cdot \left(\frac{n-c^{r-1}}{K} - 1 \right) \\
&= \#L \cdot \frac{n-c^{r-1}}{K} \\
&\geq \#L \cdot \frac{c^r(1-c^{-1})}{K} \\
&\geq \#L \cdot c^r \delta c \\
&\geq \#L \cdot \delta n.
\end{aligned}$$

Comparing it with (4.3.40), we obtain that on $A_r \cap B_r$, $\max_{l \in L} T_n(l) > \delta n$ for all $n \in [c^r, c^{r+1}]$. In particular, the set $\{\tilde{\mu}_n(l) : T_n(l) \geq \delta n, l \in L\}$ is nonempty, so

$$\begin{aligned}
\max\left\{\tilde{\mu}_n(k) : T_n(k) \geq \delta n, k \notin L\right\} &\leq \max_{j \notin L} \mu(\theta_j) + \varepsilon < \mu(\theta^*) - \varepsilon \\
&\leq \min\left\{\tilde{\mu}_n(l) : T_n(l) \geq \delta n, l \in L\right\},
\end{aligned}$$

and ϕ draws from $j_n \in L$ for $c^r \leq n \leq c^{r+1}$ on $A_r \cap B_r$.

For the final claim, by the previous argument, $[j_n \notin L] \subseteq A_r^c \cup B_r^c$, so

$$P_\theta[j_n \notin L] \leq P_\theta(A_r^c) + P_\theta(B_r^c) = o(c^{-r}),$$

for $n \in [c^r, c^{r+1}]$, so

$$\sum_{n=c^r}^{c^{r+1}} P_\theta[j_n \notin L] \leq (c^{r+1} - c^r + 1)o(c^{-r}) = c^r o(c^{-r}) + o(1) = o(1).$$

For any $N \in \mathbb{N}$, $\sum_{n=1}^{N} P_\theta[j_n \notin L] = \sum_{n=1}^{c} P_\theta[j_n \notin L] + \sum_{n=c+1}^{N} P_\theta[j_n \notin L]$. We pick a $k \in \mathbb{N}$ such that $c^k < N \leq c^{k+1}$. Then $k \leq \frac{\ln N}{\ln c}$, hence

$$\begin{aligned}
\sum_{n=c}^{N} P_\theta[j_n \notin L] &\leq \sum_{n=c}^{c^{k+1}} P_\theta[j_n \notin L] = \sum_{n=c}^{c^2} P_\theta[j_n \notin L] + \ldots + \sum_{n=c^k}^{c^{k+1}} P_\theta[j_n \notin L] \\
&= ko(1) \leq \frac{\ln N}{\ln c} o(1) = o(\ln N).
\end{aligned}$$

Combining the above together,

$$\sum_{n=1}^{N} P_\theta[j_n \notin L] = \sum_{n=1}^{c} P_\theta[j_n \notin L] + o(\ln N) = c_0 + o(\ln N) = o(\ln N),$$

as we wanted. ■

4.4 *The upper confidence bound algorithm*

A serious drawback when implementing the Lai-Robbins algorithm, is that it does not provide us with a general rule of constructing the upper confidence bounds functions g_{ni}. Designing such functions has only been achieved for specific distributions, but even when we do have examples of g_{ni}, actually computing them is demanding.

Peter Auer, Nicolò Cesa-Bianchi and Paul Fischer [ACBF02], proposed an algorithm called the *Upped Confidence Bound Algorithm* (UCB) which addresses these issues. It relies on a similar sampling rule as with the Lai-Robbins, where the populations are drawn after comparing their mean and an upper confidence bound estimates. Here the mean estimates are just the sample means and the upper confidence bounds are

$$\overline{x}_{j,n_j} + \sqrt{\frac{3\ln n}{2n_j}}, \tag{4.4.1}$$

where $\overline{x}_{j,n_j}$ is the average reward obtained so far from machine j and n_j is the number of times machine j was played during the first n rounds.[9] In essence, the principle of optimism under uncertainty is still present, but the hard task of finding and computing the Lai-Robbins functions g_{ni}'s is replaced by computing the simple expression (4.4.1). Their algorithm achieves logarithmic regret, although with a larger logarithmic constant than the Lai-Robins method.

The Upper Confidence Bound Algorithm	
Step 1	Play each machine once.
Step 2	During the $n+1$-th round, play the machine that maximizes $\overline{x}_{j,n_j} + \sqrt{\frac{3\ln n}{2n_j}}$, where $\overline{x}_{j,n_j}$ is the average reward obtained so far from machine j and n_j is the number of times machine j was played during the first n rounds.

Table 7: The Upper Confidence Bound Algorithm (UCB) [ACBF02].

Remark 4.4.1: In any implementation of the UCB algorithm, each hand is played infinitely often with probability one. This follows by an argument similar to the one used in the proof of the consistency of the Robbins rule (Theorem 4.2.4): Suppose that there exist some arms, say $A \subseteq \{1, \dots, K\}$, which after an implementation have been chosen only finitely many times. Let also $N_A < \infty$ denote the last round during which an arm from A has been chosen.

[9] The actual the upper bound formula that was given by Auer et al. was $\overline{x}_{j,n_j} + \sqrt{\frac{2\ln n}{n_j}}$, but we will use this slight modification [Mun14] as it leads to a better logarithmic bound ($6/\Delta_j^2$ compared to $8/\Delta_j^2$ of the original one). However, both of them are larger than the Lai-Robbins constant.

For every $j \notin A$, the set $N_j = \{n : \phi(n) = j\}$ is infinite, so by the Strong Law of Large Numbers, $\tilde{x}_{j,n_j} \to \mu(j)$ as $n_j \to \infty$. Since we draw from j infinitely many times, it has to be that for any $k \in A$,

$$\overline{x}_{j,n_j} + \sqrt{\tfrac{3\ln n}{2n_j}} > \overline{x}_{k,n_k} + \sqrt{\tfrac{3\ln n}{2n_k}} \geq \overline{x}_{k,n_k} + \sqrt{\tfrac{3\ln n}{2n_A}}$$

for infinitely many n's. After a few manipulations, this implies that for n large enough so that the quantity $\overline{x}_{k,n_k} - \overline{x}_{j,n_j}$ is near its limit α, we have that

$$\sqrt{\tfrac{3\ln n}{2}}\left(\tfrac{1}{\sqrt{n_j}} - \tfrac{1}{\sqrt{n_A}}\right) \geq \alpha - 1$$

holds for infinitely many n's, which is a contradiction since the LHS of this inequality tends to $-\infty$.

Theorem 4.4.2 ([ACBF02, Theorem 1])**:** *Suppose that we run the Upper Confidence Bound (UCB) Algorithm on K arms with expectations $\mu_1, \ldots, \mu_K$ and reward distributions $P_1, \ldots, P_K$ supported on $[0,1]$. Then, during the first n rounds, every suboptimal arm k is expected to be played $E[T_k(n)]$ times with*

$$E[T_k(n)] \leq \frac{6\ln n}{\Delta_k^2} + \frac{\pi^2}{3} + 1, \tag{4.4.2}$$

where $\Delta_k = \mu^* - \mu_k$. Additionally, the expected regret R_n after n rounds is at most

$$R_n := n\mu^* - \sum_{j=1}^{K} \mu_j \mathbb{E}[T_j(n)] \leq 6 \sum_{k:\Delta_k>0} \frac{\ln n}{\Delta_k} + K\left(\frac{\pi^2}{3} + 1\right). \tag{4.4.3}$$

Proof. Suppose that a suboptimal arm k is pulled during round $n > K$, namely $\phi(n) = k \notin L$. This means that

$$\overline{x}_{k,T_k(n-1)} + \sqrt{\frac{3\ln n}{2T_k(n-1)}} \geq \overline{x}_{k^*,T_{k^*}(n-1)} + \sqrt{\frac{3\ln n}{2T_{k^*}(n-1)}} \tag{4.4.4}$$

for every optimal arm $k^* \in L$. Additionally, for every $k^* \in L$, at least one of the following three claims must hold:

(i) $\overline{x}_{k^*,T_{k^*}(n-1)} + \sqrt{\frac{3\ln n}{2T_{k^*}(n-1)}} < \mu^*$,

(ii) $\overline{x}_{k,T_k(n-1)} > \mu_k + \sqrt{\frac{3\ln n}{2T_k(n-1)}}$,

(iii) $\mu_k + 2\sqrt{\frac{3\ln n}{2T_k(n-1)}} \geq \mu^*$.

Indeed, if both (i) and (ii) failed, then

$$\mu^* \leq \overline{x}_{k^*,T_{k^*}(n-1)} + \sqrt{\frac{3\ln n}{2T_{k^*}(n-1)}} \qquad \text{due to (i)}$$

$$\leq \overline{x}_{k,T_k(n-1)} + \sqrt{\frac{3\ln n}{2T_k(n-1)}} \qquad \text{due to (4.4.4)}$$

$$< \mu_k + 2\sqrt{\frac{3\ln n}{2T_k(n-1)}} \qquad \text{due to (ii)}$$

and (iii) would hold. Therefore, at every stage $n > K$, either a suboptimal is *not* pulled, or at least one of (i)-(iii) holds. Let $n > K$ and set $a_n = \frac{6\ln n}{\Delta_k^2} + 1$. Then,

$$\begin{aligned} T_k(n) = \sum_{t=1}^{n} I_{[\phi(t)=k]} &\leq \sum_{t=1}^{[a_n]} 1 + \sum_{t=[a_n]+1}^{n} I_{[\phi(t)=k, T_k(t)>[a_n]]} \\ &\leq a_n + \sum_{t=[a_n]+1}^{n} I_{[\phi(t)=k, T_k(t)>[a_n]]}. \end{aligned}$$

Let $t \geq [a_n]+1$ and consider the events $A_t = [\phi(t) = k, T_k(t) > [a_n]]$. Since

$$T_k(t) > [a_n] = \left[\frac{6\ln n}{\Delta_k^2} + 1\right] > \frac{6\ln n}{\Delta_k^2},$$

claim (iii) fails to hold[10] for any such $t = [a_n]+1, \ldots$, and since $\phi(t)$ is suboptimal, for these t's either (i) or (ii) must hold. Using Hoeffding's Inequality (Theorem B.1.7), we can bound the probability of the event appearing in (i) as follows:

$$\begin{aligned} P\left[\overline{x}_{k^*,T_{k^*}(n-1)} + \sqrt{\tfrac{3\ln n}{2T_{k^*}(n-1)}} < \mu^*\right] &\leq P\left[\overline{x}_{k^*,s} + \sqrt{\tfrac{3\ln n}{2s}} < \mu^* \text{ for some } 1 \leq s \leq t\right] \\ &\leq \sum_{s=1}^{t} P\left[\overline{x}_{k^*,s} + \sqrt{\tfrac{3\ln n}{2s}} < \mu^*\right] \\ &\leq \sum_{s=1}^{t} e^{-6\ln t} = \sum_{s=1}^{t} \frac{1}{t^6} \leq \frac{1}{t^2}. \end{aligned}$$

With a similar argument we obtain the same bound for the events in (ii), so $T_k(n) \leq a_n + \sum_{t=[a_n]+1}^{n} I_{A_t}$ with $P(A_t) \leq \frac{2}{t^2}$ for all t. Therefore,

$$\mathbb{E}[T_k(n)] \leq a_n + 2\sum_{t=[a_n]+1}^{n} \frac{2}{t^2} \leq \frac{6\ln n}{\Delta_k^2} + \frac{\pi^2}{3} + 1.$$

The bound for the expected regret is then

$$\begin{aligned} R_n = \sum_{k:\Delta_k>0}^{K} \Delta_k \mathbb{E}[T_k(n)] &\leq \sum_{k=1}^{K} \Delta_k \left(\frac{6\ln n}{\Delta_k^2} + \frac{\pi^2}{3} + 1\right) \\ &\leq 6\sum_{k:\Delta_k>0} \frac{\ln n}{\Delta_k} + K\left(\frac{\pi^2}{3} + 1\right), \end{aligned}$$

which completes the proof. ■

[10] When (iii) is true, it has to be that $T_k(t-1) \leq \frac{6\ln t}{\Delta_k^2}$.

Although the UCB algorithm achieves logarithmic regret, the logarithmic constant in the expression

$$\mathbb{E}_{\mathrm{UCB}}[T_j(n)] \leq \frac{6}{\Delta_j} \ln n + c \tag{4.4.5}$$

is greater than the one in the Lai-Robbins algorithm,

$$\mathbb{E}_{\mathrm{LR}} \leq \left(\frac{1}{I(\theta_j, \theta^*)} + o(1) \right) \ln n \leq \left(\frac{1}{2\Delta_j^2} + o(1) \right) \ln n, \tag{4.4.6}$$

since the factor $\frac{1}{I(\theta_j,\theta^*)}$ is less or equal than $\frac{1}{2\Delta_j^2}$ by the Pinsker Inequality.

The Lai-Robbins Algorithm	The UCB Algorithm
Makes certain assumptions on the reward distributions (they need to be parametric, to satisfy certain properties with respect to the Kullback-Leibler divergence, etc.).	The only assumption is that the reward distributions need to be supported in $[0,1]$.
Involves a family of "upper confidence bounds" functions g_{ni}, the construction of which is not straightforward. Even when they are available, their actual computation is usually demanding.	The "upper confidence bound" functions are simple, easily implemented and computationally efficient.
Achieves logarithmic regret asymptotically.	Achieves logarithmic regret both asymptotically and uniformly in time.
The logarithmic constant in the expression for regret is $\frac{1}{2\Delta_j}$ for any suboptimal machine j.	The corresponding regret constant is $\frac{6}{\Delta_j} > \frac{1}{2\Delta_j}$.

Table 8: A comparison between the Lai-Robbins and the Upper Confidence Bound Algorithms.

Theorem 4.4.3 (Pinsker's Inequality): [BLM13, Theorem 4.19] *Let P, Q be two probability distributions on $(X, \mathcal{A})$. Then*

$$\delta(P,Q) \leq \sqrt{\frac{1}{2} I(P,Q)}, \tag{4.4.7}$$

where $\delta(P,Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$ is the total variation distance and $I(P,Q)$ is the Kullback-Leibler divergence between the two distributions.

Corollary 4.4.4: *Let X^*, X_j be random variables on $(X,,P)$ with expectations μ^* and μ_j respectively. Set $\Delta_j = \mu^* - \mu_j$. Then*

$$I(P_{X^*}, P_{X_j}) \geq 2\Delta_j^2.$$

Proof. By Pinsker's Inequality, $2\delta(P_{X^*}, P_{X_j})^2 \leq I(P_{X^*}, P_{X_j})$, so it suffices to show that $\delta(P_{X^*}, P_{X_j}) \geq \Delta_j$:

$$\begin{aligned}\Delta_j &= \mathbb{E}[X^*] - \mathbb{E}[X_j] \\ &= \int_0^1 \big(P[X^* > t] - P[X_j > t]\big)\, dt \\ &\leq \int_0^1 \sup_A \big|P_{X^*}(A) - P_{X_j}(A)\big|\, dt \\ &= \delta(P_{X^*}, P_{X_j}),\end{aligned}$$

as we wanted. ■

Before closing this section, it is worth to mention that using the Cauchy-Schwarz Inequality, we can also obtain a bound for the regret which does not involve the unknown quantities Δ_k:

Corollary 4.4.5: *The expected regret of the UCB algorithm is bounded by*

$$R_n \leq \sqrt{Kn\left(6\ln n + \frac{\pi^2}{3} + 1\right)}. \tag{4.4.8}$$

Proof. Applying the Cauchy-Schwarz Inequality to the formula of the regret, we obtain that

$$\begin{aligned}E[R_n] = \sum_{k=1}^K \Delta_k \sqrt{\mathbb{E}[T_k(n)]}\sqrt{\mathbb{E}[T_k(n)]} &\leq \left(\sum_{k=1}^K \Delta_k^2 \mathbb{E}[T_k(n)]\right)^{1/2} \left(\sum_{k=1}^K \mathbb{E}[T_k(n)]\right)^{1/2} \\ &\leq \sqrt{n}\sqrt{\sum_{k=1}^K \Delta_k^2\left(\frac{6\ln n}{\Delta_k^2} + \frac{\pi^2}{3} + 1\right)} \leq \sqrt{nK\left(6\ln n + \left(\frac{\pi^2}{3} + 1\right)\right)},\end{aligned}$$

as $\Delta_k \in [0,1]$ for all k and thus $\sum_{k=1}^K \Delta_k^2 \leq K$. ■

4.5 *The upper confidence bound algorithm with epochs*

As we saw in the previous paragraph, the UCB algorithm achieves logarithmic regret growth, but the logarithmic constant is larger than the optimal one. In the same article, Auer et al. constructed an extension in which the logarithmic constant can get arbitrarily close to it. The main principles of this algorithm are the same as in UCB, but there are also two main differences.

Firstly, the upper confidence bounds have a slightly more involved expression, and secondly, the arm chosen in each round is not played just once, but instead the algorithm sticks with it for a whole time interval called *epoch*. The lengths of these epochs get increasingly larger the more often an arm is picked, resulting in larger exploitation progressively. The rate in which these lengths increase needs to be chosen carefully, and is closely related to the exploration-exploitation trade off.

THE UPPER CONFIDENCE BOUND ALGORITHM WITH EPOCHS	
STEP 1	Initialize with $r_j = 0$ for $j = 1, \dots, K$ and play each machine once.
STEP 2	Select the machine j that maximizes the quantity $\overline{x}_j + a_{n,r_j}$, where $\overline{x}_j$ is the average reward obtained so far by the j machine, $a_{n,r_j} = \sqrt{\frac{(1+a)\ln\frac{en}{\tau(r_j)}}{2\tau(r_j)}}$, $\tau(r) = \lceil (1+a)^r \rceil$ and n is the number of total pulls done so far overall.
STEP 3	Play the j machine exactly $\tau(r_j+1) - \tau(r_j)$ times.
STEP 4	Set $r_j := r_j + 1$ and return to STEP 2.

Table 9: The Upper Confidence Algorithm with epochs (UCB2) [ACBF02].

The algorithm contains a parameter $a \in (0,1)$ which the researcher is free to choose as he desires. The value of it has a direct effect on the rate of growth of the epoch lengths, and thus to the asymptotic behavior of the regret. Indeed, the r-th epoch of an arm when parameter a is used, has a length of $L_a(r) = \lceil (1+a)^{r+1} \rceil - \lceil (1+a)^r \rceil$. We will discuss the practical implications of this formula in the end of the paragraph, after we have established the corresponding asymptotic properties of the regret.

Remark 4.5.1: For small values of a, the quantity $\tau(r_j+1) - \tau(r_j)$ which appears in Step 3 may be equal to zero, which means that the algorithm picks a machine but never actually draws from it. However, this does not cause much trouble [Sta20] since by proceeding to Step 3, we set $r_j := r_j + 1$ and return to Step 1 for the next iteration.

As no value was drawn, all the quantities $\overline{x}_k + a_{n,r_k}$ remained unchanged for $k \neq j$. Additionally, for the j machine, we have that r_j is equal to $r_j = r + 1$ for some r with the property that $\tau(r+1) = \tau(r)$. Plugging this into a_{n,r_j}, we obtain that

$$a_{n,r_j} = a_{n,r+1} = \sqrt{\frac{(1+a)\ln\frac{en}{\tau(r+1)}}{2\tau(r+1)}} = \sqrt{\frac{(1+a)\ln\frac{en}{\tau(r)}}{2\tau(r)}} = a_{n,r},$$

so the value of a_{n,r_j} remained the same as well.

This implies that when we repeat the Step 1 of the algorithm, the same machine j will be picked again. But this time it will be played a total of $\tau(r+2) - \tau(r+1)$ times. If this number also happens to be equal to zero, then after a finite number of iterations we will definitely end up with a non-zero quantity. This is clear from the inequality $\tau(r+1) - \tau(r) \geq (1+a)^r(1+a) - 1 \to \infty$, so after a finite number of possibly "empty loops", the algorithm always pulls the arms it chooses.

Theorem 4.5.2 ([ACBF02, Theorem 2])**:** *Let $a \in (0,1)$ and suppose that the upper confidence bound algorithm with epochs (UCB2) is run on K machines having*

expectations $\mu(1),\dots,\mu(K)$ *and reward distributions* $P_1,\dots,P_K$ *respectively, supported in* $[0,1]$. *Then, the expected regret* R_n *after* $n \geq \max_{i:\mu_i<\mu^*} \frac{e}{2\Delta_i^2}$ *is at most*

$$R_n \leq \sum_{i:\mu_i<\mu^*} \left(\frac{(1+a)(1+4a)\ln(2e\Delta_i^2 n)}{2\Delta_i} + \frac{c_a}{\Delta_i} \right), \tag{4.5.1}$$

where c_a *is a constant which converges to infinity as* $a \to 0$.

Proof. Suppose that $n \geq \frac{e}{2\Delta_j^2}$ for all suboptimal arms j and let $\tilde{r}_j$ be the largest integer such that

$$\tau(\overline{r}_j - 1) \leq \frac{(1+4a)\ln(2e\Delta_j^2)}{2\Delta_j^2}.$$

For $r = 1$, we have that $\tau(r-1) = \tau(0) = 1$, whereas

$$\frac{(1+4a)\ln(2e\Delta_j^2)}{2\Delta_j^2} \geq \frac{1+4a}{\Delta_j^2} \geq 1 + 4a > 1 = \tau(r),$$

so it must be that $\tilde{r}_j \geq 1$. Due to the presence of epochs, the total number of pulls from an arm j during the first n rounds $T_j(n)$ has a more complicated expression:

$$\begin{aligned} T_j(n) &\leq 1 + \sum_{r\geq 1} (\tau(r) - \tau(r-1))\, I_{[\text{machine } j \text{ finished its } r\text{-epoch}]} \\ &= 1 + \tau(\overline{r}_j) - \tau(0) + \sum_{r\geq \overline{r}_j+1} (\tau(r) - \tau(r-1))\, I_{[\text{machine } j \text{ finished its } r\text{-epoch}]} \\ &= \tau(\overline{r}_j) + \sum_{r\geq \overline{r}_j+1} (\tau(r) - \tau(r-1))\, I_{[\text{machine } j \text{ finished its } r\text{-epoch}]}. \end{aligned} \tag{4.5.2}$$

In particular, if machine j just finished its r-epoch when the total draws were n, then $T_j(n) = 1 + \sum_{k=1}^{r} (\tau(k) - \tau(k-1)) = \tau(r)$.

Pick a suboptimal machine j and suppose that it just finished its r-epoch. Then, there must exist a previous point in time, say t, where j was picked to start its r-th epoch. At this time t, (a) the machine had already finished its $r-1$-epoch, thus had been played exactly $\tau(r-1)$-times up to that point, and (b) some epoch just finished and we entered the comparison stage. The result of the comparison was to pick machine j, thus we had that

$$\overline{x}_{j,\tau(r-1)} + a_{t,r-1} \geq \overline{x}^*_{\tau(i)} + a_{t,i}, \tag{4.5.3}$$

where $\overline{x}^*_{\tau(i)} + a_{t,i}$ was the upper confidence bound of the optimal machine. The latter machine had been chosen during i-epochs until time t for some $i \geq 0$, resulting to a total of $\tau(i)$ draws, hence the indices that appear in $\overline{x}^*_{\tau(i)} + a_{t,i}$. Notice also that $t \geq \tau(i) + \tau(r-1)$ since the optimal machine and machine j have been played $\tau(i)$ and $\tau(r-1)$ times respectively up until round t.

Therefore, there exists some $i \geq 0$ and $t \geq \tau(r-1)+\tau(i)$ such that (4.5.3) holds. In turn, this implies that either there exists some $t \geq \tau(r-1)$ with

$$\overline{x}_{j,\tau(r-1)} + a_{t,r-1} \geq \mu^* - \frac{a\Delta_j}{2},$$

or that there exists an $i \geq 0$ and a $\tau' \geq \tau(r-1)+\tau(i)$ with

$$\overline{x}^*_{\tau(i)} + a_{t',i} \leq \mu^* - \frac{a\Delta_j}{2}.$$

Indeed, otherwise for every $i \geq 0$ and $t \geq \tau(t-1)+\tau(i)$ we would have that

$$\overline{x}_{j,\tau(r-1)} + a_{t,r-1} < \mu^* - \frac{a\Delta_j}{2} < \overline{x}^*_{\tau(i)} + a_{t',i},$$

which contradicts (4.5.3).

Suppose now that machine j just finished its r-epoch during round n. Since $a_{t,r}$ is increasing with respect to t, we also have that either

$$\overline{x}_{j,\tau(r-1)} + a_{n,r-1} \geq \mu^* - \frac{a\Delta_j}{2} \tag{4.5.4}$$

or that there exists an $i \geq 0$ such that

$$\overline{x}^*_{\tau(i)} + a_{\tau(r-1)+\tau(i),i} \leq \mu^* - \frac{a\Delta_j}{2}. \tag{4.5.5}$$

By taking expectations in (4.5.2),

$$\begin{aligned}
\mathbb{E}[T_j(n)] &\leq \tau(\overline{r}_j) + \sum_{r \geq \overline{r}_j+1} (\tau(r)-\tau(r-1))\, P\left[\text{machine } j \text{ finished its } r\text{-epoch}\right] \\
&\leq \tau(\overline{r}_j) + \sum_{r \geq \overline{r}_j+1} (\tau(r)-\tau(r-1))\, P\left[\overline{x}_{j,\tau(r-1)} + a_{n,r-1} \geq \mu^* - \tfrac{a\Delta_j}{2}\right] \\
&\quad + \sum_{r \geq \overline{r}_j+1} \sum_{i \geq 1} (\tau(r)-\tau(r-1))\, P\left[\overline{x}^*_{\tau(i)} + a_{\tau(r-1)+\tau(i),i} \leq \mu^* - \tfrac{a\Delta_j}{2}\right]
\end{aligned}$$

and we need to bound the probabilities of these two events. Firstly, observe that by the definition of $\overline{r}_j$, for every $r > \overline{r}_j$, we have that $\tau(r-1) > \frac{(1+4a)\ln(2e\Delta_j^2)}{2\Delta_j^2} > \frac{1}{2\Delta_j^2}$, so

$$a_{n,r-1} = \sqrt{\frac{(1+a)\ln\frac{en}{\tau(r-1)}}{2\tau(r-1)}} \leq \sqrt{\frac{(1+a)\ln\frac{en}{\tau(r-1)}}{\frac{(1+4a)\ln(2e\Delta_j^2)}{\Delta_j^2}}} \leq \Delta_j\sqrt{\frac{1+a}{1+4a}}, \tag{4.5.6}$$

which is also less or equal than $\Delta_j(1+a)$ for $a < 0.1$. We bound the probability of the events $\left[\overline{x}_{j,\tau(r-1)} + a_{n,r-1} \geq \mu^* - \frac{a\Delta_j}{2}\right]$ using Hoeffding's Inequality. The random variables appearing in the Cesàro sum $\overline{x}_{j,\tau(r-1)}$ are i.i.d. with mean μ_j, so

$$P\left[\overline{x}_{j,\tau(r-1)} + a_{n,r-1} \geq \mu^* - \tfrac{a\Delta_j}{2}\right] = P\left[\overline{x}_{j,\tau(r-1)} + a_{n,r-1} \geq \mu_j + \Delta_j - \tfrac{a\Delta_j}{2}\right]$$

$$
\begin{aligned}
&= P\left[\overline{x}_{j,\tau(r-1)} \geq \mu_j + \frac{\left(\Delta_j - \frac{a\Delta_j}{2} - a_{n,r-1}\right)\tau(r-1)}{\tau(r-1)}\right] \\
&\leq \exp\left(-2\tau(r-1)\left(\Delta_j - \frac{a\Delta_j}{2} - a_{n,r-1}\right)^2\right) \\
&\leq \exp\left(-2\tau(r-1)\Delta_j^2\left(1 - \frac{a}{2} - (1-a)\right)^2\right) \\
&\leq \exp\left(-\frac{\tau(r-1)\Delta_j^2 a^2}{2}\right) \qquad (4.5.7)
\end{aligned}
$$

for $a < 0.1$. Using the elementary inequality

$$
\begin{aligned}
\tau(r) = \lceil (1+a)^r \rceil &\leq (1+a)^r + 1 = (1+a)^{r-1}(1+a) + 1 \\
&\leq \lceil (1+a)^{r-1} \rceil (1+a) + 1 = \tau(r-1)(1+a) + 1,
\end{aligned}
$$

we have that for the function $g(x) = \frac{x-1}{1+a}$, the inequality $g(x) \leq \tau(r-1)$ holds for every $r \geq 1$ and $x \in [\tau(r-1), \tau(r)]$. Set $c = \Delta_j^2 a^2 < 1$. Then,

$$
\begin{aligned}
\int_0^\infty e^{-cg(x)}dx &= \sum_{r=1}^{\infty}\int_{\tau(r-1)}^{\tau(r)} e^{-cg(x)}dx \\
&\geq \sum_{r=1}^{\infty}\int_{\tau(r-1)}^{\tau(r)} e^{-c\tau(r-1)}dx \\
&= \sum_{r=1}^{\infty}(\tau(r-1) - \tau(r))\, e^{-c\tau(r-1)} \\
&\geq \sum_{r \geq \overline{r}_j + 1}(\tau(r-1) - \tau(r))\, P\left[\overline{x}_{j,\tau(r-1)} + a_{n,r-1} \geq \mu^* - \frac{a\Delta_j}{2}\right].
\end{aligned}
$$

Additionally, $\int_0^\infty e^{-cg(x)}dx = e^{\frac{c}{1+a}}\frac{1+a}{c} \leq \frac{(1+a)e}{\Delta_j^2 a^2}$, so

$$
\sum_{r \geq \overline{r}_j + 1}(\tau(r-1) - \tau(r))\, P\left[\overline{x}_{j,\tau(r-1)} + a_{n,r-1} \geq \mu^* - \frac{a\Delta_j}{2}\right] \leq \frac{(1+a)e}{\Delta_j^2 a^2}. \quad (4.5.8)
$$

For the events $\left[\overline{x}^*_{\tau(i)} + a_{\tau(r-1)+\tau(i),i} \leq \mu^* - \frac{a\Delta_j}{2}\right]$, using again Hoeffding's Inequality, we have that

$$
P\left[\overline{x}^*_{\tau(i)} + a_{\tau(r-1)+\tau(i),i} \leq \mu^* - \frac{a\Delta_j}{2}\right] \leq \exp\left(-\tau(i)\frac{a^2\Delta_j^2}{2} - (1+a)\ln\left(e^{\frac{\tau(r-1)+\tau(i)}{\tau(i)}}\right)\right)
$$

and the expression

$$
A = \sum_{r \geq \overline{r}_j + 1}\sum_{i \geq 1}(\tau(r) - \tau(r-1))\, P\left[\overline{x}^*_{\tau(i)} + a_{\tau(r-1)+\tau(i),i} \leq \mu^* - \frac{a\Delta_j}{2}\right]
$$

can be bounded by

$$
A \leq \sum_{i \geq 0}\exp\left(-\frac{\tau(i)a^2\Delta_j^2}{2}\right)\sum_{r \geq \overline{r}_j + 1}(\tau(r) - \tau(r-1))\left(1 + \frac{\tau(r-1)}{\tau(i)}\right)^{-1-a}.
$$

We define the function $h(x) = \left(1 + \frac{x-1}{(1+a)\tau(i)}\right)^{-1-a}$. Then $h(x) \geq \left(1 + \frac{\tau(r-1)}{\tau(i)}\right)^{-1-a}$ for every $x \in [\tau(r-1), \tau(r)]$ and for all i,

$$\begin{aligned}\tau(i)\left(\frac{1+a}{a}\right)^{1+a} &\geq \tau(i)\frac{1+a}{a}\left(1 - \frac{1}{(1+a)\tau(i)}\right)^{-a} = \int_0^\infty g(x)dx \\ &\geq \sum_{r=1}^{\infty}\int_{\tau(r-1)}^{\tau(r)}\left(1 + \tfrac{\tau(r-1)}{\tau(i)}\right)^{-1-a}dx \\ &= \sum_{r=1}^{\infty}(\tau(r) - \tau(r-1))\left(1 + \tfrac{\tau(r-1)}{\tau(i)}\right)^{-1-a},\end{aligned}$$

so

$$A \leq \left(\frac{1+a}{a}\right)^{1+a}\sum_{i\geq 0}\tau(i)\exp\left(-\frac{\tau(i)a^2\Delta_j^2}{2}\right). \tag{4.5.9}$$

We define the function $w(x) = (1+a)^x + 1$. Then, for every $x \in [i, i+1]$, we have that $w(x) \geq \tau(i) \geq (1+a)^{x-1}$, so

$$\begin{aligned}1 + \int_1^\infty w(x)\exp\left(-\frac{(1+a)^{x-1}a^2\Delta_j^2}{2}\right)dx &= 1 + \sum_{i=1}^{\infty}\int_i^{i+1} w(x)\exp\left(-\frac{(1+a)^{x-1}a^2\Delta_j^2}{2}\right)dx \\ &\geq 1 + \sum_{i=1}^{\infty}\int_i^{i+1}\tau(i)\exp\left(-\frac{(1+a)^{x-1}a^2\Delta_j^2}{2}\right)dx \\ &\geq 1 + \sum_{i=1}^{\infty}\int_i^{i+1}\tau(i)\exp\left(-\frac{\tau(i)a^2\Delta_j^2}{2}\right)dx \\ &= \sum_{i=0}^{\infty}\tau(i)\exp\left(-\frac{\tau(i)a^2\Delta_j^2}{2}\right).\end{aligned}$$

One can verify that

$$1 + \int_1^\infty w(x)\exp\left(-\frac{(1+a)^{x-1}a^2\Delta_j^2}{2}\right)dx \leq 1 + \frac{1}{\ln(1+a)}\left[\frac{e^{-\lambda}}{\lambda} + \int_\lambda^\infty \frac{e^{-x}}{x}dx\right],$$

where $\lambda = \frac{a^2\Delta_j^2}{2(1+a)} < \frac{1}{4}$. The last expression $F(\lambda) = \frac{e^{-\lambda}}{\lambda} + \int_\lambda^\infty \frac{e^{-x}}{x}dx$ can be bounded from above by $F(\lambda) \leq \frac{11}{10\lambda} = \frac{11(1+a)}{5a^2\Delta_j^2}$. Combining all the previous steps, we obtain that

$$\mathbb{E}[T_j(n)] \leq \tau(\overline{r}_j) + \frac{(1+a)e}{\Delta_j^2 a^2} + \left(\frac{1+a}{a}\right)^{1+a}\left(1 + \frac{1}{\ln(1+a)}\cdot\frac{11(1+a)}{5a^2\Delta_j^2}\right),$$

with $\tau(\overline{r}_j) \leq (1+a)\tau(\overline{r}_j - 1) + 1 \leq \frac{(1+a)(1+4a)\ln(2en\Delta_j^2)}{2\Delta_j^2} + 1$, therefore,

$$\begin{aligned}\mathbb{E}[T_j(n)] &\leq \tfrac{(1+a)(1+4a)\ln(2en\Delta_j^2)}{2\Delta_j^2} + 1 + \tfrac{(1+a)e}{\Delta_j^2 a^2} + \left(\tfrac{1+a}{a}\right)^{1+a}\left(1 + \tfrac{1}{\ln(1+a)}\cdot\tfrac{11(1+a)}{5a^2\Delta_j^2}\right) \\ &\leq \tfrac{(1+a)(1+4a)\ln(2en\Delta_j^2)}{2\Delta_j^2} + \tfrac{1}{\Delta_j^2}\left(\tfrac{(1+a)e}{a^2} + \left(\tfrac{1+a}{a}\right)^{1+a}\left(1 + \tfrac{11(1+a)}{5a^2\ln(1+a)}\right)\right)\end{aligned}$$

$$= \frac{(1+a)(1+4a)\ln n}{2\Delta_j^2} + \frac{c'_a}{\Delta_j^2}, \tag{4.5.10}$$

for $c'_a = \frac{(1+a)(1+4a)\ln(2e\Delta_j^2)}{2\Delta_j^2} + 1 + \frac{(1+a)e}{a^2} + \left(\frac{1+a}{a}\right)^{1+a}\left(1 + \frac{11(1+a)}{5a^2\ln(1+a)}\right)$. ■

Relation (4.5.10) is particularly informative regarding the algorithm's behavior. Its first term reveals that, regardless of the choice for $a \in (0,1)$, the algorithm always leads us to a logarithmic asymptotic regret. The second term is a constant that we may not avoid, and despite the fact that its effect will vanish as $n \to \infty$, it may not be negligible for small n's. By picking a sufficiently small, we can approach the optimal asymptotic lower bound arbitrarily close. However, although the first term $\frac{(1+a)(1+4a)\ln n}{2\Delta_j^2}$ converges to the Lai-Robbins bound $\frac{\ln n}{2\Delta_j^2}$ as $a \to 0$, the corresponding second term c'_a tends to infinity. This means that the closer we want to get to the optimal bound asymptotically, the higher the toll we have to pay in finite time.

This is a typical aspect of the exploitation-exploration trade off that we mentioned in the beginning on the paragraph, when we stressed out the significance of the rate in which the epoch lengths increase. By definition, the length $L_a(r)$ of the r-th epoch of an arm is equal to $L_a(r) = \lceil (1+a)^{r+1} \rceil - \lceil (1+a)^r \rceil$, and one can easily verify that the sequence $(L_a(r))_r$ increases more slowly as a decreases. This has the implication that for small values of a, the first epochs last relatively short, giving the algorithm more opportunities to explore between the arms. This increased tendency to explore, especially in early stages, leads to the appearance of the large constant c'_a.

References

[ACBF02] P. Auer, N. Cesa-Bianchi, P. Fischer, Finite-time analysis of the multiarmed bandit problem, *Machine Learning* **47**, pp. 235–256, 2002. DOI: 10.1023/A:1013689704352 Cited on p. xiv, 15, 106, 107, 111

[BLM13] S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence,* Oxford University Press, 2013. ISBN: 9780199535255 Cited on p. 89, 109, 129, 130

[CR75] Y. Chow, T. Lai, Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings, *Trans. Amer. Math. Soc.* **208**, pp. 51-72, 1975. DOI: 10.1090/S0002-9947-1975-0380973-6 Cited on p. 98

[Doo90] J. Doob, *Stochastic Processes*, Wiley Classics Library Ed., 1990. ISBN: 978-0-471-52369-7 Cited on p. 98

[Gut13] A. Gut, *Probability: A Graduate Course*, Springer Texts in Statistics 75, Springer-Verlag New York, 2013. DOI: 10.1007/978-1-4614-4708-5 Cited on p. 98, 131, 132, 133

[Joy11] J. Joyce, Kullback-Leibler Divergence, In: Lovric M. (eds) International Encyclopedia of Statistical Science, Springer 2011. DOI: 10.1007/978-3-642-04898-2_327 Cited on p. 89

[LR85] T. Lai, H. Robbins, Asymptotically efficient adaptive allocation rules, *Advances in Applied Mathematics* **6** (1), pp. 4-22, 1985. DOI: 10.1016/0196-8858(85)90002-8 Cited on p. xiv, 13, 89, 93, 97, 100, 101

[LS20] T. Lattimore, C Szepesvári, *Bandit Algorithms,* Cambridge University Press, 2020. ISBN: 9781108571401, Cited on p.

[MZ37] J. Marcinkiewicz, A. Zygmund, Sur les fonctions indépendantes, *Fundamenta Mathematicae* **29** (1), pp. 60-90, 1937. URL: http://eudml.org/doc/212925 Cited on p. 98

[Mun14] R. Munos, From Bandits to Monte-Carlo Tree Search: The Optimistic Principle Applied to Optimization and Planning, *Foundations and Trends in Machine Learning* **7** (1) 2014. URL: 10.1561/2200000038 Cited on p. 106

[Rob52] H. Robbins, Some aspects of the sequential design of experiments, *Bull. Amer. Math. Soc.* **58** (5), pp. 527-535, 1952. DOI: S0002-9904-1952-09620-8 Cited on p. xiv, 13, 86, 88

[Saa17] Saad, *A generalization of Kolmogorov's maximal inequality,* URL: https://math.stackexchange.com/q/2566297 Cited on p. 98

[Sta20] N. Stamatis, *Calclulation of epoch length in UCB2,* URL: https://stats.stackexchange.com/q/448406 Cited on p. 111

[Wai19a] M. Wainwright, *High-Dimensional Statistics, A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019. ISBN: 9781108627771 Cited on p. 89

Part III

APPENDIX

In our thesis we heavily relied on results from diverse fields of Pure Mathematics, most of which are not usually taught at an undergraduate level. In this appendix, we present them in a relatively brief but coherent manner. The curious reader will certainly find fascinating the fact that in order to establish results with such important practical applications, it was necessary to build on deep theoretical tools.

A

Functional Analysis

This appendix contains basic definitions and results that we need from Functional Analysis. We use [Arg04] and [Gia03] as our main references, although complete treatments and proofs of them can be found in any Functional Analysis or Hilbert Space Theory book.

A.1 *Metric spaces*

We first mention a few results from metric spaces that we need.

Definition A.1.1: Let X be a nonempty set. A metric ρ on X is a function $\rho : X \times X \to \mathbb{R}$ that satisfies the following properties:

(i) For every $x, y \in X$, $\rho(x, y) \geq 0$ and $\rho(x, y) = 0$ if and only if $x = y$.

(ii) For every $x, y \in X$, $\rho(x, y) = \rho(y, x)$.

(iii) For every $x, y, z \in X$, $\rho(x, y) \leq \rho(x, z) + \rho(y, z)$.

The pair (X, ρ) is called a *metric space.*

A metric space (X, ρ) is called *complete* if every Cauchy sequence is convergent. In any metric space, if we consider a finite family of open and dense sets $(U_i)_{i=1}^n$, their intersection $U = \bigcap_{i=1}^n U_i$ is also open and dense. This result does not continue to hold in general when the family $(U_i)_{i=1}^\infty$ is countable. Baire's Category Theorem asserts that this intersection is dense, if one works in a complete metric space.

Theorem A.1.2 (Baire): *Let (X, ρ) be a complete metric space and $(U_i)_{i=1}^\infty$ be a sequence of open and dense subsets of it. Then their intersection $U = \bigcap_{i=1}^\infty U_i$ is dense in X.*

A very useful consequence of Baire's Theorem is the fact that if a complete space can be written as the countable union of closed sets, then at least one of them has a nonempty interior.

Theorem A.1.3: *Let (X, ρ) be a complete metric space and $(F_n)_n$ be a sequence of closed subsets of it with $X = \bigcup F_n$. Then there exists some $n_0 \in \mathbb{N}$ such that $\mathring{F}_{n_0} \neq \emptyset$.*

A.2 *Normed spaces*

Definition A.2.1: A *normed space* is a pair $(X, \|\cdot\|)$, where X is a real vector space, and $\|\cdot\| : X \to \mathbb{R}$ is a real function that satisfies the following axioms:

(i) For every $x \in X$, $\|x\| \geq 0$, and $\|x\| = 0$ if and only if $x = 0$.

(ii) For every $x \in X$ and $\lambda \in \mathbb{R}$, $\|\lambda x\| = \lambda \|x\|$.

(iii) For every $x, y \in X$, $\|x + y\| \leq \|x\| + \|y\|$.

A complete normed space is called a *Banach space*.

Every norm induces a metric ρ on X, by defining $\rho(x, y) = \|x - y\|$ for every $x, y \in X$. In this way, tools and results from real analysis can be used, or even be strengthened significantly, during the study of normed spaces.

Definition A.2.2: Let $(X, \|\cdot\|)$ be a normed space and $f : X \to \mathbb{R}$ be a real function on X. We say that f is *linear* if $f(\lambda x + \mu y) = \lambda f(x) + \mu f(y)$ for every $x, y \in X$ and $\lambda, \mu \in \mathbb{R}$. If, in addition, f is continuous, it is called a *bounded linear functional*, or simply a bounded functional.[1]

The space X^* is always a Banach space, under the norm defined by

$$\|x^*\| = \sup_{x \in B_X} |x^*(x)|, \quad \text{for } x^* \in X^*,$$

where $B_X = \{x \in X| : \|x\| \leq 1\}$ denotes the unit ball of X. Although X^* is always nonempty, as it contains the zero functional, it is highly nontrivial to show that it contains nonzero elements when X is infinite-dimensional. This is asserted by the celebrated Hahn-Banach theorem, along with its consequences [Arg04, Chapter 5], which show that X^* has a wealth of elements.[2]

Theorem A.2.3 (Hahn-Banach): *Let X be a linear space and $p : X \to \mathbb{R}$ be a sublinear functional, namely $p(\lambda x) = \lambda p(x)$ and $p(x+y) \leq p(x)+p(y)$ for every $\lambda > 0$ and $x, y \in X$. Suppose that Z is a linear subspace of X and $f : Z \to \mathbb{R}$ is a linear functional with the property that $f(x) \leq p(x)$ for every $x \in X$. Then there exists some linear functional $\tilde{f} : X \to \mathbb{R}$ such that $\tilde{f}(x) = f(x)$ for every $x \in Z$ and $\tilde{f}(x) \leq p(x)$ for every $x \in X$.*

The property that $\tilde{f}(x) = f(x)$ on Z denotes that $\tilde{f}$ is an extension of f, whereas the property that $\tilde{f}(x) \leq p(x)$ for every $x \in X$ denotes that $\tilde{f}$ is still dominated by p on the whole space, and not only on Z. The most common proofs of the Hahn-Banach theorem require some form of the Axiom of Choice, although there exists proofs without the use of it.

One of the most important applications of the Hahn-Banach theorem, is that one can always separate any element x of a normed space X from any closed subspace of it $Y \subseteq X$ which does not contain x, with a bounded and linear functional. This consequence was used in the proof of Cybenko's Universal Approximation Theorem:

[1] The term bounded comes from the property that a linear functional f is continuous if and only if $|f(x)| \leq M\|x\|$ for some $M > 0$ for every $x \in X$. Clearly, nontrivial linear functionals can never be bounded in the usual sense, since their image is the whole real line $f(X) = \mathbb{R}$, however the term bounded has prevailed due to the aforementioned property.

[2] In cardinality terms, X^* contains at least as many elements as X [KS16]. This is equivalent to the fact that $\dim X^* \geq \dim X$.

Proposition A.2.4: *Let $(X, \|\cdot\|)$ be a normed space, $Y \subseteq X$ be a closed subspace of it and let $x \in X \setminus Y$. Then there exists some $f \in X^*$ such that $f(x) = d(x, Y) > 0$, $\|f\| = 1$ and $f(y) = 0$ for every $y \in Y$.*

A.3 *Hilbert spaces*

Definition A.3.1: Let X be a vector space. A mapping $\langle\cdot,\cdot\rangle : X \times X \to \mathbb{R}$ is called an *inner product* if it satisfies the following axioms:

(i) For every $x \in X$, $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$.

(ii) For every $x, y \in X$, $\langle x, y\rangle = \langle y, x\rangle$.

(iii) For every $x, y, z \in X$ and $\lambda, \mu \in \mathbb{R}$, $\langle \lambda x + \mu z, y\rangle = \lambda\langle x, y\rangle + \mu\langle z, y\rangle$.

In every inner product space, the Cauchy-Schwarz inequality holds:

Theorem A.3.2 (Cauchy-Schwarz inequality): *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. Then for every $x, y \in X$,*

$$|\langle x, y\rangle| \leq \sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}, \tag{A.3.1}$$

with equality if and only if x and y are linearly dependent.

Every inner product induces a norm on the underlying space, defined by $\|x\| = \sqrt{\langle x, x\rangle}$. If X is complete under this norm, then X is called a *Hilbert space*. An easy consequence of the Cauchy-Schwarz inequality is that the inner product $\langle\cdot,\cdot\rangle : X \times X \to \mathbb{R}$ is a continuous function when $X \times X$ is equipped with any product[3] metric [Arg04, Proposition 4.4]:

Proposition A.3.3: *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. If $x_n \to x$ and $y_n \to y$, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$.*

Norms induced by inner products, enjoy two very useful geometric properties, the Parallelogram Law[4] and the Pythagorean Theorem. Their proofs follow immediately from the defining properties of the inner product:

Proposition A.3.4 (Parallelogram Law): *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. For every $x, y \in X$, the parallelogram identity holds*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2. \tag{A.3.2}$$

[3] Things get more complicated if one asks whether the inner product is a uniformly continuous function. In this case, the answer depends on which product metric has been used and, not surprisingly, for every inner product space $(X, \langle\cdot,\cdot\rangle)$, there always exist two distinct, but equivalent, product metrics on $X \times X$, such that $\langle\cdot,\cdot\rangle$ is uniformly continuous with respect to one but not with respect to the other.

[4] As a matter of fact, inner product spaces not only have property (A.3.2), but are completely characterized by it, in the sense that a norm is induced by an inner product, if and only if it satisfies the Parallelogram Law. In order to prove the converse direction, one considers a norm $\|\cdot\|$ which satisfies the Parallelogram Law, and defines a function $\phi : X \times X \to \mathbb{R}$ as $\phi(x, y) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2)$. With some moderate effort, one can show that ϕ is an inner product which induces $\|\cdot\|$.

Definition A.3.5: Two elements x, y of an inner product space are said to be *orthogonal*, if $\langle x, y\rangle = 0$. We usually use the symbol $x \perp y$ to denote this. Similarly, if A, B are subsets of X, they are called *orthogonal* when $a \perp b$ for every $a \in A$ and $b \in B$.

Proposition A.3.6 (Pythagorean Theorem): *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space and $x, y \in X$ such that $\langle x, y\rangle = 0$. Then*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2. \tag{A.3.3}$$

If an inner product space is also a Banach space under the norm induced by the inner product, then it is called a Hilbert space.

Definition A.3.7: A complete inner product space is called a *Hilbert space.*

The typical example of a Hilbert space is

$$\ell_2(\mathbb{N}) = \left\{(a_n)_n : \sum_n a_n^2 < \infty\right\},$$

the space of square summable real sequences,[5] equipped with the inner product defined by $\langle x, y\rangle = \sum_{n=1}^{\infty} x_n y_n$ for every $x = (x_n)_n$ and $y = (y_n)_n$ in ℓ_2. In a similar manner, one can define the ℓ_p space for every $1 \le p < \infty$ as the space that contains all sequences $(a_n)_n$ such that $\sum_{n=1}^{\infty} |a_n|^p < \infty$. All the ℓ_p spaces are Banach spaces, however ℓ_2 is the only Hilbert space among them.[6] The basic inequality $|a|^q \le |a|^p$ which holds for every $p \le q$ and a such that $|a| < 1$, and the fact that any summable sequence has to converge to zero, thus eventually the absolute value of its terms are less than one, has as a consequence that $\ell_p \subseteq \ell_q$ for every $p \le q$. This inclusion is always strict when $p < q$. In particular, we have that $\ell_1 \subset \ell_2$, and in the spirit of Robbins-Monro, we say that the elements in the set $\ell_2 \setminus \ell_1$ are *type* $1/n$ *sequences.*

In our thesis we are mostly working on the Hilbert space of square integrable functions on some compact set $X \subseteq \mathbb{R}^d$:

$$L_2(X) = \left\{f : X \to \mathbb{R} : \int_X f(x)^2 d\lambda(x) < \infty\right\},$$

equipped with the inner product

$$\langle f, g\rangle = \int_X f(x)g(x)d\lambda(x)$$

for $f, g \in L_2(X)$.

Definition A.3.8: A family $\{e_i : i \in I\}$ in an inner product space is called *orthonormal*, if $\|e_i\| = 1$ for every $i \in I$ and $e_i \perp e_j$ for every $i \neq j \in I$.

[5] In a certain sense, ℓ_2 is the only Hilbert space as every Hilbert space H is isometrically isomorphic to $\ell_2(\kappa)$ for some set κ of cardinality equal to the Hilbert dimension of H. By the term *Hilbert dimension* we mean the cardinality of the smallest orthonormal basis of H.

[6] For any $p \neq 2$, one can easily find $x, y \in \ell_p$ for which the Parallelogram Law fails to hold.

It is easy to see that every orthonormal family is a linearly independent set. If $\{e_1, \dots, e_n\}$ is a finite orthonormal family and F_n is the subspace generated by it, then $\{e_1, \dots, e_n\}$ is called an *orthonormal basis* of F_n. Furthermore, for every finite n-dimensional subspace F of an inner product space, we can find an orthonormal basis of it, having cardinality n, using the Gram-Schmidt process.

Perhaps one of the most striking results in Hilbert spaces H, that may not necessarily hold in an arbitrary normed space, is the fact that for every $x \in H$ and every closed subspace M, there always exists a unique projection of x on M. Recall that if x is an element of a metric space (X, ρ) and $M \subseteq X$ is a subset of it, we define the distance between x and M as $d(x, M) = \inf\{\rho(x, m) : m \in M\}$. In general this infimum may not be attained, even when M is a closed set. [7]

However, when $(X, \|\cdot\|)$ is a Hilbert space and M is a closed and convex subset of it, there always exists an $m_0 \in M$ such that $d(x, M) = \rho(x, m_0) = \|x - m_0\|$. Furthermore, m_0 is unique and has the property that $x - m_0$ is orthogonal to M. This result also applies when M is a closed subspace of X, as every subspace is a convex set. We will state the result for finite dimensional subspaces F, although it holds for infinite dimensional subspaces as well, with some obvious modifications [Arg04, Proposition 4.21]:

Proposition A.3.9: *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space and F be a finite dimensional subspace of it, having $\{e_1, \dots, e_n\}$ as an orthonormal basis. If $x \in X$, then the closest point of x to F is the element $y_0 = \sum_{i=1}^{n} \langle x, e_i\rangle e_i$, that is, y_0 satisfies the property that $\|x - y_0\| = d(x, F)$. Additionally, $x - y_0$ is orthogonal to F.*

The element y_0 of the previous proposition is called the *projection* of x on the subspace F, and is usually denoted by $y_0 = P_F(x)$. The projection of x on F is the only element z of F with the property that $x - z$ is orthogonal to F:

Proposition A.3.10: *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space, F be a closed subspace of it and $P_F(x)$ be the projection of x on F. Suppose that $z \in F$ satisfies the property that $x - z$ is orthogonal to F. Then $z = P_F(x)$.*

Proof. We have that

$$\begin{aligned}\|z - P_F(x)\|^2 &= \langle z - x + x - P_F(x), z - P_F(x)\rangle \\ &= \langle z - x, z - P_F(x)\rangle + \langle x - P_F(x), z - P_F(x)\rangle \\ &= 0,\end{aligned}$$

both $z - x$, $x - P_F(x)$ being orthogonal to F, and $z - P_F(x) \in F$. ∎

A.4 *Topological vector spaces*

In every normed space $(X, \|\cdot\|)$ the addition and scalar multiplication operations are continuous. Topological vector spaces are the natural generalization of

[7] As an example, one can take $X = (0, 1) \cup \{2\}$ equipped with the metric induced by the usual metric in $\mathbb{R}$. Then $F = (0, 1)$ is a closed set in X, the distance between 2 and F is equal to one, yet it is not attained.

normed spaces, where the norm is replaced by a topology τ that "respects" the linear structure of X, meaning that the vector space operations are continuous with respect to τ. To avoid trivialities, we usually work with Hausdorff[8] topologies.

Definition A.4.1: A vector space X equipped with a Hausdorff topology τ such that the addition operation $+ : X \times X \to X$ and the scalar multiplication operation $\cdot : \mathbb{R} \times X \to X$ are continuous, is called a *topological vector space*.

Definition A.4.2: A topological vector space (X, τ) is called *locally convex* if every neighborhood of zero contains a convex neighborhood of zero. A completely metrizable[9] locally convex space, is called a *Fréchet space*.

Since Fréchet spaces are complete, Baire's theorem is applicable. Thus, if we write a Fréchet (X, τ) as the countable union of closed sets, $X = \bigcup_{n=1}^{\infty} F_n$, at least one of them must have a nonempty interior. If, in addition, each F_n is a subspace of X, then there exists a subspace F_{n_0} of X with a nonempty interior, which yields that $X = F_{n_0}$. This fact was used in the proof of Lemma 2.3.9, p. 34.

Definition A.4.3: Let (X, τ) be a topological vector space. A set $A \subseteq X$ is called *absorbing* if for every $x \in X$, there exists some $\lambda_0 > 0$ such that $\lambda x \in A$ for every $0 \leq \lambda \leq \lambda_0$.

Theorem A.4.4: *If a subspace Y of a topological vector space X has a nonempty interior, then $Y = X$.*

Proof. Any topological vector space contains a neighborhood base consisting of absorbing sets [AB06, Structure Theorem 5.6]. Suppose that Y is a subspace of X with a nonempty interior. This means that, there exists some $y_0 \in Y$ and an open set U such that $y_0 \in U$. Then $0 \in U - y_0$ which is also an open set contained in Y, since Y is a subspace. Let W be an absorbing open set contained in $U - y_0$. Then $W \subseteq Y$ and for every $x \in X$, $\lambda x \in W \subseteq Y$ for some $\lambda > 0$, thus $X \subseteq \cup_{\lambda > 0} \lambda Y = Y$. ∎

[8] A topological space (X, τ) is said to be a Hausdorff space, or to satisfy the T_2 axiom, if for every $x, y \in X$ with $x \neq y$, there exist two disjoint open sets U_x and U_y, such that $x \in U_x$ and $y \in U_y$.

[9] A topological space (X, τ) is called completely metrizable, if there exists some metric ρ on X, such that (X, ρ) is a complete metric space and ρ induces the topology τ on X.

B

Probability Theory

B.1 *Subgaussian random variables*

Subgaussian random variables have the defining property that their tail probabilities are bounded by the respective probabilities of a gaussian random variable. Because of this property, they are only candidate random variables for which the Hoeffding inequality can hold. Most of the results mentioned in this section can be found in Omar Rivasplata's expository article [Riv12].

Definition B.1.1: A random variable X is said to be *σ-subgaussian*, if for every $t \in \mathbb{R}$,

$$M_X(t) := \mathbb{E}\left[e^{tX}\right] \leq e^{\sigma^2 t^2/2}. \tag{B.1.1}$$

The right hand side of (B.1.1) is just the moment generating function of a normal random variable with mean zero and variance equal to σ^2. So, a random variable is σ-subgaussian when its moment generating function is dominated by the one of $N(0, \sigma^2)$.

Subgaussian random variables have always zero mean and a finite variance, bounded by, but not necessarily equal to, the σ^2 that appears in (B.1.1).

Proposition B.1.2: *If a random variable X is b-subgaussian, then $\mathbb{E}[X] = 0$ and $V(X) \leq b^2$.*

Proof. Using the Taylor's expansion and Lebesgue's Dominated Convergence Theorem,

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{t^n X^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbb{E}\left[X^n\right] = 1 + t\mathbb{E}[X] + \frac{t^2}{2}\mathbb{E}\left[X^2\right] + o(t^2) \leq e^{\frac{b^2 t^2}{2}}$$

as $t \to 0$. Dividing with t and taking limits as $t \to 0$, we obtain that $\mathbb{E}[X] \leq 0$. Repeating the same argument for $-X$ gives us the opposite inequality.

Dividing with t^2 yields that $\mathbb{E}[X^2] = V(X) \leq b^2$. In fact, using the formula $\mathbb{E}|X^p| = \int_0^\infty P[|X|^p \geq t]dt$ and the tail bound property that we will mention shortly, one can show that every moment of a subgaussian r.v. X is finite. ∎

By abusing the definition and the previous proposition, we may use the term subgaussian even for random variables X for which $\mathbb{E}[X] \neq 0$, implying that $Y = X - \mathbb{E}[X]$ is subgaussian.

The sum of two subgaussian random variables is also subgaussian. Additionally, scalar products of subgaussian random variables remain subgaussian:

Theorem B.1.3: *Let $(\Omega, \mathcal{A}, P)$ be a measurable space and let G denote the set of all subgaussian random variables on it. Then G is a vector space under the usual addition and scalar product.*

Proof. Suppose that X, Y are b_1 and b_2-subgaussian random variables respectively and let $c \in \mathbb{R}$.

For $t \in \mathbb{R}$,

$$M_{cX}(t) = \mathbb{E}\left[e^{tcX}\right] = M_X(ct) \leq e^{\frac{t^2c^2b^2}{2}},$$

so $cX \in G$, and in particular, it is $|c|b$-subgaussian.

Regarding their sum, by Hölder's Inequality we have that for every $p, q > 1$ with $\frac{1}{p} + \frac{1}{q}$,

$$\begin{aligned}\mathbb{E}\left[e^{t(X+Y)}\right] &\leq \mathbb{E}\left[e^{tpX}\right]^{1/p} \mathbb{E}\left[e^{tqY}\right]^{1/q} \leq \left(\exp\frac{t^2p^2b_1^2}{2}\right)^{1/p}\left(\exp\frac{t^2q^2b_2^2}{2}\right)^{1/q} \\ &= e^{\frac{t^2}{2}(pb_1^2+qb_2^2)}.\end{aligned}$$

The function $f(p) = pb_1^2 + qb_2^2 = pb_1^2 + \frac{p}{1-p}b_2^2$ attains its minimum for $p_0 = \frac{b_1+b_2}{b_1}$, with $f(p_0) = (b_1+b_2)^2$, therefore $X+Y$ is (b_1+b_2)-subgaussian. ∎

A simple application of the Markov inequality is that for a b-subgaussian random variable X, its tails satisfy the properties

$$P(X \geq \lambda) \leq e^{-\frac{\lambda^2}{2b^2}} \text{ and } \quad P(X \leq -\lambda) \leq e^{-\frac{\lambda^2}{2b^2}} \tag{B.1.2}$$

for every $\lambda > 0$. Indeed, for every $t > 0$,

$$P(X \geq \lambda) = P(tX \geq t\lambda) \leq \frac{\mathbb{E}\left[e^{tX}\right]}{e^{\lambda t}} \leq e^{-\lambda t + \frac{b^2t^2}{2}}$$

and for $t^* = \frac{\lambda}{b^2}$ we obtain the desired upper bound. Applying the result to $-X$, yields the respective lower tail bound.

In fact, inequalities (B.1.2) completely characterize subgaussian random variables. We collect all the equivalent definitions below.

Theorem B.1.4: [Ver18, Proposition 2.5.2] *Let X be a centered random variable. The following are equivalent:*

(i) There exists some $b > 0$ such that $\mathbb{E}\left[e^{tX}\right] \leq e^{\frac{b^2t^2}{2}}$ for every $t \in \mathbb{R}$.

(ii) There exists some $c > 0$ such that

$$P(X \geq \lambda) \leq e^{-c\lambda^2} \text{ and } \quad P(X \leq -\lambda) \leq e^{-c\lambda^2}$$

for all $\lambda > 0$.

(iii) There exists some $C > 0$ such that $\|X\|_p \leq C\sqrt{p}$ for every $p \geq 1$.

(iv) There exists some $a > 0$ such that $\mathbb{E}\left[e^{aX^2}\right] \leq 2$.

The typical example of a subgaussian random variable are normal random variables $N(0, \sigma^2)$ with mean zero. Another extremely useful class of examples are random variables taking values on some finite interval. Their subgaussianity is asserted by Hoeffding's Lemma.

Lemma B.1.5 (Popovicius Inequality): [ABCD05] *Let X be a random variable that takes values in the interval $[a, b]$ with probability one. Then $V(X) \leq \frac{(b-a)^2}{4}$.*

Proof. The variance of X has the property that $V(X) = \min_c \mathbb{E}(X-c)^2$. For $c_0 = \frac{a+b}{2}$, we notice that $|X - \frac{a+b}{2}| \leq \frac{b-a}{2}$, so $V(X) \leq \mathbb{E}\left|X - \frac{a+b}{2}\right|^2 \leq \frac{(b-a)^2}{4}$. ∎

Lemma B.1.6 (Hoeffding): [BLM13, Lemma 2.2] *A centered random variable $X : (\Omega, \mathcal{A}, P) \to \mathbb{R}$, taking values on $[a, b]$ a.s., is $\frac{b-a}{2}$-subgaussian.*

Proof. For every $\lambda > 0$, we define the random variable $Y_\lambda : \Omega \to \mathbb{R}$ as $Y_\lambda = \frac{e^{\lambda Y}}{M_Y(\lambda)}$, where $M_Y(\lambda) = \mathbb{E}_P\left[e^{\lambda Y}\right]$ is the m.g.f. of Y. Let also $P_\lambda = Y_\lambda \cdot P$ denote the probability measure induced by the random variable Y_λ, namely

$$P_\lambda(A) = \int_A Y_\lambda \, dP = \int_A \frac{e^{\lambda Y}}{M_Y(\lambda)} dP \quad \text{for all } A \in \mathcal{A}.$$

By Theorem 4.3.2, for every random variable Z, we also have that

$$\mathbb{E}_{P_\lambda}[Z] = \mathbb{E}_P\left[\frac{Ze^{\lambda Y}}{M_Y(\lambda)}\right].$$

In particular, setting $Z = X$ and $Z = X^2$, yields that

$$\mathbb{E}_{P_\lambda}[Y] = \mathbb{E}_P\left[\frac{Ye^{\lambda Y}}{M_Y(\lambda)}\right] \quad \text{and} \quad \mathbb{E}_{P_\lambda}\left[Y^2\right] = \mathbb{E}_P\left[\frac{Y^2 e^{\lambda Y}}{M_Y(\lambda)}\right]. \tag{B.1.3}$$

Let $\psi_Y(\lambda) = \ln \mathbb{E}_P\left[e^{\lambda Y}\right]$. Then,

$$\psi'_Y(\lambda) = \frac{\mathbb{E}\left[Ye^{\lambda Y}\right]}{M_Y(\lambda)} \quad \text{and} \quad \psi''_Y(\lambda) \frac{\mathbb{E}\left[Y^2 e^{\lambda Y}\right]}{M_Y(\lambda)} - \left(\frac{\mathbb{E}\left[Ye^{\lambda Y}\right]}{M_Y(\lambda)}\right)^2,$$

and by comparing it with (B.1.3), we obtain that

$$\psi''_Y(\lambda) = \mathbb{E}_{P_\lambda}\left[Y^2\right] - \left(\mathbb{E}_{P_\lambda}[Y]\right)^2 = V_{P_\lambda}[Y_\lambda] \leq \frac{(b-a)^2}{4},$$

by Popovicius Inequality, since $P_\lambda[Y \in [a, b]] = 1$. Fix some $\lambda > 0$. By Taylor's theorem, there exists some $\theta \in [0, \lambda]$ such that

$$\psi_Y(\lambda) = \psi_Y(0) + \lambda \psi'_Y(0) + \frac{\lambda^2}{2}\psi''_Y(\theta) = \frac{\lambda^2}{2}\psi''_Y(\theta) \leq \frac{\lambda^2 (b-a)^2}{8}.$$

The proof is concluded by taking exponents in the last inequality. ∎

Since sums of subgaussian r.v. are also subgaussian, if we begin with a sequence $(X_n)_n$ of such r.v., then each of their partial sums $S_n = X_1 + \ldots + X_n$ is also subgaussian, so one can bound the tails of these sums, either using the subgaussian property (B.1.2) directly, or with an argument similar to the one that appears in the proof of it.

Theorem B.1.7 (Hoeffding's inequality): [BLM13, Theorem 2.8] *Let $X_1, \ldots, X_n$ be independent random variables such that each X_i takes its values in $[a_i, b_i]$ almost surely. Let S_n be their centered partial sum, $S_n = \sum_{i=1}^n (X_i - \mathbb{E}[X_i])$. Then for every $t > 0$,*

$$P(S_n \geq t) \leq e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}} \quad \text{and} \quad P(S_n \leq -t) \leq e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}. \tag{B.1.4}$$

Proof. Each of the random variables $Y_i = X_i - \mathbb{E}[X_i]$ is centered and takes values on $[a_i - \mathbb{E}[X_i], b_i - \mathbb{E}[X_i]]$ a.s., so it is $\frac{b_i - a_1}{2}$-subgaussian. Set $S_n = Y_1 + \ldots + Y_n$. By the independence of $(X_i)_i$,

$$\mathbb{E}\left[e^{\lambda S_n}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{\lambda Y_i}\right] \leq e^{\frac{\lambda^2 \sum_{i=1}^n (b_i - a_i)^2}{8}}$$

and by the Chernoff method,

$$P[S_n \geq t] = P\left[e^{\lambda S_n} \leq e^{\lambda t}\right] \leq \frac{\mathbb{E}\left[e^{\lambda S_n}\right]}{e^{\lambda t}} \leq e^{\frac{\lambda^2 \sum_{i=1}^n (b_i - a_i)^2}{8} - \lambda t}$$

for all $\lambda > 0$. The last quantity is minimized for $\lambda^* = \frac{t}{4 \sum_{i=1}^n (b_i - a_i)^2}$, resulting in (B.1.4). ∎

The vector space of the subgaussian random variables G has a richer structure. Clearly, if a random variable is σ-subgaussian, then it is also τ-subgaussian for every $\tau \geq \sigma$. This observation leads us to define the subgaussian moment of a random variable X, as the smallest constant σ for which X is σ-subgaussian:

Definition B.1.8: Let X be a subgaussian random variable. We define its *subgaussian moment* $\sigma(X)$ as

$$\sigma(X) = \inf\left\{\sigma \geq 0 : \mathbb{E}\left[e^{tX}\right] \leq e^{\sigma^2 t^2/2} \text{ for all } t \in \mathbb{R}\right\}. \tag{B.1.5}$$

This infimum is in fact a minimum: Let $(\sigma_n)_n$ be a non-increasing sequence of non-negative numbers such that $\sigma_n \to \sigma(X)$. For every $t \in \mathbb{R}$, we have that $M_X(t) \leq e^{\sigma_n^2 t^2/2} \to e^{\sigma(X)^2 t^2/2}$, so the number $\sigma(X)$ also belongs to the set appearing in the right hand side of (B.1.5), which proves that every subgaussian random variable X is $\sigma(X)$-subgaussian.

The function σ defines a complete norm on the space of subgaussian random variables.

Theorem B.1.9: *The space (G, σ) is a normed space.*

Proof. If $X = 0$ a.s., then $\mathbb{E}\left[e^{\lambda X}\right] = 1 \leq e^{b^2 t^2/2}$ for every $b \geq 0$ and $t \in \mathbb{R}$, so $\sigma(X) = 0$. Conversely, suppose that X is such that $\sigma(X) = 0$. Then, for any fixed $\lambda \in \mathbb{R}$, we have that $\mathbb{E}\left[e^{\lambda X}\right] \leq 1$. The function $f(x) = e^{\lambda x}$ is convex, so by the Jensen inequality, $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$, namely $\mathbb{E}\left[e^{\lambda X}\right] \geq e^0 = 1$, which implies that $M_X(\lambda) = 1$ for every $\lambda \in \mathbb{R}$. But this is just the m.g.f. of the random variable Y which is zero almost surely. By the uniqueness of the m.g.f. [Gut13, Theorem 8.1] we conclude that X and Y have the same distribution, so $X = 0$ almost surely.

We now prove that σ is positively homogenous. Let $X \in G$ and $a \in \mathbb{R}$. We observe that

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}} \ \forall \lambda \in \mathbb{R} \iff \mathbb{E}\left[e^{a\lambda X}\right] \leq e^{\frac{a^2\lambda^2\sigma^2}{2}} \ \forall \lambda \in \mathbb{R},$$

so

$$\begin{aligned}
\sigma(X) &= \inf\left\{\sigma > 0 : \mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}}\right\} = \inf\left\{\sigma > 0 : \mathbb{E}\left[e^{\lambda aX}\right] \leq e^{\frac{\lambda^2(a\sigma)^2}{2}}\right\} \\
&= \frac{1}{|a|}\inf\left\{|a|\sigma > 0 : \mathbb{E}\left[e^{\lambda aX}\right] \leq e^{\frac{\lambda^2(a\sigma)^2}{2}}\right\} \\
&= \frac{1}{|a|}\inf\left\{\tau > 0 : \mathbb{E}\left[e^{\lambda aX}\right] \leq e^{\frac{\lambda^2\tau^2}{2}}\right\} \\
&= \frac{1}{|a|}\sigma(aX),
\end{aligned}$$

thus $\sigma(aX) = |a|\sigma(X)$.

Lastly, let $X, Y \in G$. Since the infima in $\sigma(X)$ and $\sigma(Y)$ are attained, X and Y are $\sigma(X)$ and $\sigma(Y)$-subgaussians respectively. As we saw in the proof of Theorem B.1.3, $X + Y$ is $\sigma(X) + \sigma(Y)$-subgaussian, so $\sigma(X + Y) \leq \sigma(X) + \sigma(Y)$, and σ is subadditive. ∎

Theorem B.1.10: [BK80, Theorem 1]. *The space (G, σ) is a Banach space.*

Proof. Let $(X_n)_n$ be a Cauchy sequence in G under the σ norm. We observe that for every $n, m \in \mathbb{N}$,

$$V(X_n - X_m) = \mathbb{E}\left[(X_n - X_m)^2\right] \leq \sigma(X_n - X_m),$$

so $(X_n)_n$ is also Cauchy in L_2, thus convergent to some $X \in L_2$. By the inequality $\|Z\|_1 \leq \|Z\|_2$ which holds for every $Z \in L_2$ [AB06, Corollary 13.3], we also have that $X \in L_1$ and $X_n \to X$ in L_1 as well. By [AB06, Theorem 13.6], we can pass to a subsequence of it, say $(X_n)_{n \in N_1}$, which converges to X almost surely.

We will show that $X \in G$ and also that $\sigma(X_n - X) \to 0$. For the first claim, notice that for every $\lambda \in \mathbb{R}$ and $\varepsilon > 0$,

$$\sup_{n \in N_1} \mathbb{E}\left[\left(e^{\lambda X_n}\right)^{1+\varepsilon}\right] = \sup_{n \in N_1} \mathbb{E}\left[e^{\lambda(1+\varepsilon)X_n}\right] \leq \sup_{n \in N_1} e^{\frac{\lambda^2(1+\varepsilon)^2\sigma(X_n)^2}{2}} < \infty,$$

the last supremum being finite because the sequence $(\sigma(X_n))_n$ is bounded.[1] By [Gut13, Theorem 4.2], the sequence $(e^{\lambda X_n})_{n\in N_1}$ is uniformly integrable and it also converges to $e^{\lambda X}$ almost surely. By Theorem B.2.4,

$$\mathbb{E}\left[e^{\lambda X}\right] = \lim_{n\in N_1} \mathbb{E}\left[e^{\lambda X_n}\right] \leq \lim_{n\in N_1} e^{\frac{\lambda^2\sigma^2(X_n)}{2}} \leq e^{\frac{\lambda^2 \sup_n \sigma^2(X_n)}{2}},$$

which implies that X is subgaussian with $\sigma(X) \leq \sup_n \sigma(X_n)$.

For every n, the random variable $X - X_n$ is subgaussian and the sequence of random variables $(\exp(\lambda(X_m - X_n)))_n$ us uniformly integrable. Indeed, for every $\lambda \in \mathbb{R}$ and $\varepsilon > 0$,

$$\begin{aligned}\sup_{m\in N_1, m\geq n} \mathbb{E}\left[e^{\lambda(1+\varepsilon)(X_m - X_n)}\right] &\leq \sup_{m\in N_1, m\geq n} e^{\frac{\lambda^2(1+\varepsilon)^2\sigma^2(X_m - X_n)}{2}} \\ &\leq e^{\frac{\lambda^2(1+\varepsilon)^2}{2}\sup_{m\in N_1, m\geq n}\sigma^2(X_m - X_n)} < \infty.\end{aligned}$$

Additionally, $\lim_{m\in N_1, m\geq n} e^{\lambda(X_m - X_n)} = e^{\lambda(X - X_n)}$ a.s., so

$$\mathbb{E}\left[e^{\lambda(X-X_n)}\right] = \lim_{m\in N_1, m\geq n} \mathbb{E}\left[e^{\lambda(X_m - X_n)}\right] \leq e^{\frac{\lambda^2}{2}\sup_m \sigma^2(X_m - X_n)},$$

thus $\sigma(X - X_n) \leq \sup_{m\in N_1, m\geq n} \sigma^2(X_m - X_n) \to 0$ as $n \to \infty$, establishing that $(X_n)_{n\in N_1}$ converges to X in σ. Since $(X_n)_n$ is a Cauchy sequence in σ with a convergent subsequence, the original sequence $(X_n)_n$ is also convergent [Arg11, Proposition 8.2]. ■

B.2 *Convergence of random variables*

We will assume throughout that the reader is familiar with the various notions of convergence of random variables. We only mention a few results which are not always taught in probability courses. First of all, we need to recall the absolute continuity property of the Lebesgue integral [Sar18, Proposition 4.29]:

Proposition B.2.1: *Let X be an integrable random variable on $E \subseteq \mathbb{R}$. Then for every $\varepsilon > 0$, there exists some $\delta > 0$ such that $\int_A |f| dm < \varepsilon$ for every measurable A with $m(A) < \delta$.*

Definition B.2.2: A sequence of random variables $(X_n)_n$ defined on a probability space $(\Omega, \mathcal{A}, P)$ is called *uniformly integrable* if

$$\lim_{a\to\infty} \sup_n \mathbb{E}[|X_n| I_{|X_n|\leq a|}] = 0.$$

The following equivalent definition of uniform integrability is often useful [Gut13, Theorem 4.1].

Theorem B.2.3: *A sequence of random variables $(X_n)_n$ is uniformly integrable if and only if the following two conditions are met:*

[1] The inequality $|\sigma(X_n) - \sigma(X_m)| \leq \sigma(X_n - X_m)$ implies that $(\sigma(X_n))_n$ is Cauchy, thus bounded.

(i) *There exists some* $M > 0$ *such that* $\|X_n\|_1 \leq M$ *for every* $n \in \mathbb{N}$.

(ii) *For every* $\varepsilon > 0$ *there exists some* $\delta > 0$, *such that for every* A *with* $P(A) < \delta$, $\int_A |X_n| < \varepsilon$ *holds for every* $n \in \mathbb{N}$.

The notion of uniform integrability is extremely useful, as it provides us with a necessary and sufficient condition to be able to conclude the convergence of the expectations, given that a sequence converges in probability [Gut13, Theorem 5.4]:

Theorem B.2.4: *Let* $(X_n)_n$ *be a sequence of random variables such that* X_n *converges to a random variable* X *in probability. The following are equivalent*

(i) *The sequence* $(X_n)_n$ *is uniformly integrable.*

(ii) $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

B.3 *Conditional expectation*

In connection with the Hilbert Space Theory, the conditional expectation of a square integrable random variable X can be seen as the projection of X to an appropriate closed subspace of L_2. This property was used during the proof of Proposition 3.2.11.

Lemma B.3.1: *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\mathcal{G} \subseteq \mathcal{F}$ *be a sigma-algebra. Then* $L_2(\Omega, \mathcal{G}, P)$ *is a closed subspace of* $L_2(\Omega, \mathcal{F}, P)$.

Proof. If $f \in L_2(\Omega, \mathcal{G}, P)$, then f is $\mathcal{G}$-measurable, thus $\mathcal{F}$-measurable due to the inclusion $\mathcal{G} \subseteq \mathcal{F}$. Additionally, $\int_\Omega |f|^2 dP < \infty$, so $f \in L_2(\Omega, \mathcal{F}, P)$, which proves that $L_2(\Omega, \mathcal{G}, P)$ is a subset of $L_2(\Omega, \mathcal{F}, P)$. Since $L_2(\Omega, \mathcal{G}, P)$ is always a vector space, it follows that it is also a subspace of it.

To show that it is closed, let $(f_n)_n$ be a sequence in $L_2(\Omega, \mathcal{G}, P)$ that converges to $f \in L_2(\Omega, \mathcal{F}, P)$. By [AB06, Theorem 13.6], $(f_n)_n$ has a subsequence that converges to f almost surely, say $f_{k_n} \to f$. Since each f_{k_n} is $\mathcal{G}$-measurable, so is their limit f, concluding the proof. ∎

Proposition B.3.2: *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\mathcal{G} \subseteq \mathcal{F}$ *be a sigma-algebra. Suppose that* $X \in L_2(\Omega, \mathcal{F}, P)$. *Then the conditional expecttion* $\mathbb{E}[X \mid \mathcal{G}]$ *is the orthogonal projection of* X *on the closed subspace* $L_2(\Omega, \mathcal{G}, P)$.

Proof. We will show that $X - \mathbb{E}[X \mid \mathcal{G}] \perp L_2(\Omega, \mathcal{G}, P)$. Let $Z \in L_2(\Omega, \mathcal{G}, P)$. Then

$$\mathbb{E}[(Z(X - \mathbb{E}[X \mid \mathcal{G}])] = \mathbb{E}[\mathbb{E}[(Z(X - \mathbb{E}[X \mid \mathcal{G}])] \mid \mathcal{G}] = 0,$$

since $\mathbb{E}[XZ|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$ and $\mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$. By Proposition A.3.10, we have that $\mathbb{E}[X \mid \mathcal{G}] = P_{L_2(\Omega,\mathcal{G},P)}(X)$. ∎

A well known property of the conditional expectation is that

$$\mathbb{E}[XY \mid \mathcal{G}] = X\mathbb{E}[Y \mid \mathcal{G}]$$

whenever X is $\mathcal{G}$-measurable. The following proposition extends this result when the usual product of X and Y is replaced by their inner product.

Proposition B.3.3: *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$, be a sigma-algebra. Suppose that $Y \in L_2(\Omega, \mathcal{F}, P)$ and that $X \in L_2(\Omega, \mathcal{G}, P)$. Then*

$$\mathbb{E}[\langle X, Y\rangle \mid \mathcal{G}] = \langle X, \mathbb{E}[Y \mid \mathcal{G}]\rangle. \tag{B.3.1}$$

Proof. Suppose first that $X = I_A$ is a characteristic function with $A \in \mathcal{G}$. For every $B \in \mathcal{F}$,

$$\begin{aligned}
\int_B \langle X, \mathbb{E}[Y \mid \mathcal{G}]\rangle dP &= \int_B \langle I_A, \mathbb{E}[Y \mid \mathcal{G}]\rangle dP = \int_B \left(\int_\Omega I_A \mathbb{E}[Y \mid \mathcal{G}] dP \right) dP \\
&= \int_B \left(\int_A Y dP \right) dP \\
&= \int_B \langle I_A, Y\rangle dP \\
&= \int_B \langle X, Y\rangle dP \\
&= \int_B \mathbb{E}\left[\langle X, Y\rangle \mid \mathcal{G}\right] dP.
\end{aligned}$$

If $X = \sum_{i=1}^n a_i I_{A_i}$ is a simple function with $A_i \in \mathcal{G}$ for all i, then for every $B \in \mathcal{F}$,

$$\begin{aligned}
\int_B \langle X, \mathbb{E}[Y \mid \mathcal{G}]\rangle dP &= \int_B \left\langle \sum_{i=1}^n a_i I_{A_i}, \mathbb{E}[Y \mid \mathcal{G}] \right\rangle dP \\
&= \int_B \sum_{i=1}^n a_i \langle I_{A_i}, \mathbb{E}[Y \mid \mathcal{G}]\rangle dP \\
&= \sum_{i=1}^n a_i \int_B \langle I_{A_i}, \mathbb{E}[Y \mid \mathcal{F}]\rangle dP \\
&= \sum_{i=1}^n a_i \int_B \mathbb{E}[\langle I_{A_i}, Y\rangle \mid \mathcal{G}] dP \\
&= \int_B \mathbb{E}\left[\left\langle \sum_{i=1}^n a_i I_{A_i}, Y \right\rangle \right] dP \\
&= \int_B \langle X, Y\rangle dP,
\end{aligned}$$

by the previous step and the linearity of the conditional expectation.

Suppose now that $X \geq 0$ is non-negative, square-integrable and $\mathcal{G}$- measurable function. Let $(s_n)_n$ be an increasing sequence of simple functions that converges to it pointwise. Then $(\langle s_n, Y\rangle)_n$ is also increasing, bounded by $Z = \langle X, Y\rangle \in L_1$ and converges to $\langle X, Y\rangle$ pointwise. By Lebesgue's Dominated Convergence Theorem, we have that $\mathbb{E}[\langle s_n, Y\rangle \mid \mathcal{G}] \to \mathbb{E}[\langle X, Y\rangle]$. In a similar manner

$$\mathbb{E}[\langle s_n, \mathbb{E}[Y \mid \mathcal{G}]\rangle] \to \mathbb{E}[\langle X, \mathbb{E}[Y \mid \mathcal{G}\rangle] = \mathbb{E}[\langle X, Y\rangle].$$

For the general case, X we can write it as a difference of two non-negative functions $X = X^+ - X^-$, and the result follows immediately from the linearity of the conditional expectation. ■

The proofs of the following theorems can be found in [Nev75].

Theorem B.3.4 (Supermartingale Convergence Theorem): *Let $(X_n)_n$, $(Y_n)_n$ and $(Z_n)_n$ be three sequences of random variables and $(\mathcal{F}_n)_n$ be a sequence of nested sigma-algebras. Suppose that*

(a) Each X_n, Y_n, Z_n is non-negative and $\mathcal{F}_n$-measurable.

(b) For every $n \in \mathbb{N}$, we have that $\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \leq Y_n - X_n + Z_n$.

(c) The series $\sum_{n=0}^{\infty} Z_n$ converges.

Then the series $\sum_{n=0}^{\infty} X_n$ also converges and there exists some non-negative random variable Y, such that $Y_n \to Y$ almost surely.

Theorem B.3.5 (Martingale Convergence Theorem): *Let $(X_n)_n$ be a sequence of random variables and $(\mathcal{F}_n)_n$ be a sequence of nested sigma-algebras. Suppose that*

(a) Each X_n is $\mathcal{F}_n$-measurable.

(b) For every $n \in \mathbb{N}$, we have that $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$.

(c) There exists a constant $M > 0$ such that $\mathbb{E}[|X_n|] \leq M$ for every $n \in \mathbb{N}$.

Then $(X_n)_n$ converges almost surely to a random variable X.

C

Additional Topics

C.1 *Natural density*

When trying to compare two infinite subsets of $\mathbb{N}$ we often face the paradox that although one may intuitively seem to be "larger" than the other, they both have the same cardinality. For example, one can consider the sets $A = \mathbb{N}$ and $B = 2\mathbb{N}$. Both A and B are countably infinite, but as one goes through $\mathbb{N}$, encounters elements of A twice as often as elements of B. The notion of cardinality is not strong enough to distinguish between these two sets, and for this reason we introduce the concept of the natural density:

Definition C.1.1: Let $J \subseteq \mathbb{N}$. We define its *natural density* $d(J)$ as

$$d(J) = \lim_{n\to\infty} \frac{\#J \cap \{1,\dots,n\}}{n}, \tag{C.1.1}$$

provided that this limit exists.

The natural density d is not defined for every subset of $\mathbb{N}$ and even if we consider the family of sets $\mathcal{A}$ on which d is well defined, $\mathcal{A}$ is not an algebra [Did13]. However, d exhibits some probability measure-like properties that proved useful during the sequential design of experiments treatment:

Lemma C.1.2: *Let d denote the natural density on $\mathbb{N}$.*

(i) The density of the empty set is equal to zero, and the density of $\mathbb{N}$ is equal to one.

(ii) If A, B are two disjoint subsets of $\mathbb{N}$ for which $d(A)$ and $d(B)$ is well defined, then $d(A \cup B)$ is also well defined and $d(A \cup B) = d(A) + d(B)$.

Proof. The proof of (i) is obvious. For the second part,

$$\begin{aligned} d(A \cup B) &= \lim_{n\to\infty} \frac{\#(A \cup B) \cap \{1,\dots,n\}}{n} \\ &= \lim_{n\to\infty} \frac{\#A \cap \{1,\dots,n\}}{n} + \lim_{n\to\infty} \frac{\#B \cap \{1,\dots,n\}}{n} \\ &= d(A) + d(B), \end{aligned}$$

which proves the finite additivity of d. ∎

Although d is finitely additive, it is not σ-additive. For example, the disjoint sets $A_n = \{n\}$ all have density $d(A_n) = 0$, but their union is equal to $\mathbb{N}$, which has density one.

Bibliography

[AF03] R. Adams, J. Fournier, *Sobolev Spaces*, Academic Press, 2rd ed., 2003. ISBN: 9780120441433 Cited on p. 34

[ABCD05] R. Agarwal, N. Barnett, P. Cerone, S. Dragomir, A Survey on some Inequalities for Expectation and Variance, *Computers and Mathematics with Applications*, **49**, pp. 429-480, 2005. DOI: 10.1016/j.camwa.2002.11.026 Cited on p. 129

[AB06] C. Aliprantis, K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, 3rd ed., 2006. DOI: 10.1007/978-3-662-03004-2 Cited on p. 21, 22, 24, 25, 26, 28, 33, 40, 90, 126, 131, 133

[AT07] C. Aliprantis, R. Tourky, *Cones and Duality*, American Mathematical Society, 2007. ISBN: 978-0-8218-4146-4 Cited on p.

[Arg04] S. Argyros, *Lecture Notes in Functional Analysis* (in greek), Athens, 2004. https://semfe.gr/files/users/795/synarthsiakh_analysh_I_biblio.pdf Cited on p. 22, 35, 36, 54, 121, 122, 123, 125

[Arg11] S. Argyros, *Lecture Notes in Real Analysis* (in greek), Athens, 2011. https://semfe.gr/files/users/795/pragmatiki_book.pdf Cited on p. 25, 35, 42, 132

[Arn57] V. Arnold, On functions of three variables, *Dokl. Akad. Nauk SSSR*, **114**, pp. 679-681, 1957. DOI: 10.1007/978-3-642-01742-1_2 Cited on p. 5, 39

[ACBF02] P. Auer, N. Cesa-Bianchi, P. Fischer, Finite-time analysis of the multiarmed bandit problem, *Machine Learning* **47**, pp. 235–256, 2002. DOI: 10.1023/A:1013689704352 Cited on p. xiv, 15, 106, 107, 111

[Ban22] S. Banach, Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales, *Fundamenta Mathematicae,* **3** (1), pp. 133-181, 1922. DOI: 10.4064/fm-3-1-133-181 Cited on p. 52

[BK80] V. Buldygin, Y. Kozachenko, Subgaussian random variables, *Ukrainian Math J.,* **32**, pp. 483-489, 1980. DOI: 10.1007/BF01087176 Cited on p. 131

[BT96] D. Bertsekas, J. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996. ISBN: 1-886529-10-8 Cited on p. 11, 19, 54, 75, 80, 82

[BLM13] S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence,* Oxford University Press, 2013. ISBN: 9780199535255 Cited on p. 89, 109, 129, 130

[CCL91] T. Chen, H. Chen, R. Liu, A Constructive Proof and An Extension of Cybenko's Approximation Theorem, in: *Computing Science and Statistics, Proceedings of the 22nd Symposium on the Interface*, Springer, Berlin, pp. 163–168, 1991. DOI: 10.1007/978-1-4612-2856-1_21 Cited on p. 4, 31

[CR75] Y. Chow, T. Lai, Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings, *Trans. Amer. Math. Soc.* **208**, pp. 51-72, 1975. DOI: 10.1090/S0002-9947-1975-0380973-6 Cited on p. 98

[Cyb89] G. Cybenko, Approximation by Superposition of a Sigmoidal Function, *Mathematics of Control, Signal, and Systems*, **2**, 1989, pp. 303-314. DOI: 10.1007/BF02551274 Cited on p. 2, 21

[Did13] Did, *Sets with no asymptotical density over* $\mathbb{N}$, URL: https://math.stackexchange.com/q/492786 Cited on p. 137

[Doo90] J. Doob, *Stochastic Processes*, Wiley Classics Library Ed., 1990. ISBN: 978-0-471-52369-7 Cited on p. 98

[Fri63] A. Friedman, *Generalized Functions and Partial Differential Equations*, Prentice Hall, 1963. Cited on p. 35

[Gia03] A. Giannopoulos, *Lecture Notes in Functional Analysis* (in greek), Crete, 2003. http://users.uoa.gr/~apgiannop/fa.pdf Cited on p. 121

[GP89] F. Girosi, T. Poggio, Representation properties of networks: Kolmogorov's theorem is irrelevant, *Neural Computation*, **1**, (4), 1989. DOI: neco.1989.1.4.465 Cited on p. 6

[Gut13] A. Gut, *Probability: A Graduate Course*, Springer Texts in Statistics 75, Springer-Verlag New York, 2013. DOI: 10.1007/978-1-4614-4708-5 Cited on p. 98, 131, 132, 133

[Hec87] R. Hecht-Nielsen, Kolmogorov's mapping neural network existence theorem, in: *Proceedings of the International Conference on Neural Networks*, IEEE Press, 1987. ID: 118526925 Cited on p. 6, 40

[HSW89] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Networks,* **2** (5), pp. 359-366, 1989. DOI: 10.1016/0893-6080(89)90020-8 Cited on p. 30

[Hor91] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks,* **4** (2), pp. 251-257, 1991. DOI: 10.1016/0893-6080(91)90009-T Cited on p. 31

[Joy11] J. Joyce, Kullback-Leibler Divergence, In: Lovric M. (eds) International Encyclopedia of Statistical Science, Springer 2011. DOI: 10.1007/978-3-642-04898-2_327 Cited on p. 89

[KS16] G. Katsimpas, N. Stamatis, *Is it true that* $\dim X \leq \dim X^*$ *for every infinite dimensional Banach space X?,* URL: https://math.stackexchange.com/q/952585 Cited on p. 122

[Kol57] A. Kolmogorov, On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition, *Dokl. Akad. Nauk SSSR,* **114** (5), pp. 953–956, 1957. MR111809 Cited on p. 5, 39

[KY97] T.-Y. Kwok, D.-Y. Yeung, Objective functions for training new hidden units in constructive neural networks, *IEEE Trans Neural Netw.,* **8**(5), pp. 1131-1148, 1997. DOI: 10.1109/72.623214 Cited on p. xiv, 4, 36, 38, 39

[Kur91] V. Kurkova, Kolmogorov's theorem is relevant, *Neural Computation,* **3** (4), pp. 617-622, 1991. DOI: 10.1162/neco.1991.3.4.617 Cited on p. 6, 40

[Kur92] V. Kurkova, Kolmogorov's theorem and multilayer neural networks, *Neural Networks,* **5** (3), pp. 501-506 , 1992. DOI: 10.1016/0893-6080(92)90012-8 Cited on p. 6, 40

[LR85] T. Lai, H. Robbins, Asymptotically efficient adaptive allocation rules, *Advances in Applied Mathematics* **6** (1), pp. 4-22, 1985. DOI: 10.1016/0196-8858(85)90002-8 Cited on p. xiv, 13, 89, 93, 97, 100, 101

[LS20] T. Lattimore, C Szepesvári, *Bandit Algorithms,* Cambridge University Press, 2020. ISBN: 9781108571401, Cited on p.

[LLPS93] M. Leshno, V. Lin, A. Pinkus, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, *Neural Networks,* **6** (6), pp. 861-867, 1993. DOI: 10.1016/S0893-6080(05)80131-5 Cited on p. 3, 31, 33

[MP99] V. Maiorov, A. Pinkus, Lower Bounds for Approximation by MLP Neural Networks, *Neurocomputing,* **25**, pp. 81-91, 1999. DOI: 10.1.1.6.5161 Cited on p. 6, 7, 40

[Mun13] J. Munkres, *Topology,* Pearson New International Ed., 2013. Cited on p. 26

[Mun14] R. Munos, From Bandits to Monte-Carlo Tree Search: The Optimistic Principle Applied to Optimization and Planning, *Foundations and Trends in Machine Learning* **7** (1) 2014. URL: 10.1561/2200000038 Cited on p. 106

[Nev75] J. Neveu, *Discrete Parameter Martingales*, North-Holland, 1975. ISBN: 978-0-7204-2810-0 Cited on p. 135

[Pin99] A. Pinkus, Approximation theory of the MLP model in neural networks, *Acta Numerica,* **8**, pp. 143-195, 1999. DOI: 10.1017/S0962492900002919 Cited on p. 32

[Riv12] O. Rivasplata, Subgaussian random variables: An expository note, 2012. DOI: 10.13140/RG.2.2.36288.23040 Cited on p. 127

[Rob52] H. Robbins, Some aspects of the sequential design of experiments, *Bull. Amer. Math. Soc.* **58** (5), pp. 527-535, 1952. DOI: S0002-9904-1952-09620-8 Cited on p. xiv, 13, 86, 88

[RM51] H. Robbins, S. Monro, A Stochastic Approximation Method, *Ann. Math. Statist.* **22** (3), pp. 400-407, 1951. DOI: 10.1214/aoms/1177729586 Cited on p. xiv, 7, 47, 48, 52

[Ros83] S. Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press, 1983. ISBN: 9781483269092 Cited on p. 73

[RF10] H. Royden, P. Fitzpatrick, *Real Analysis,* Prentice Hall, 4th Ed., 2010. Cited on p. 27

[Saa17] Saad, *A generalization of Kolmogorov's maximal inequality,* URL: https://math.stackexchange.com/q/2566297 Cited on p. 98

[Sar18] I. Sarantopoulos, *An introduction to Measure and Integration Theory* (in greek), Athens, 2018. http://mycourses.ntua.gr/courses/SEMFE1028/document/IntroMTI.pdf Cited on p. 132

[Sta20] N. Stamatis, *Calclulation of epoch length in UCB2,* URL: https://stats.stackexchange.com/q/448406 Cited on p. 111

[Tsi94] J. Tsitsiklis, Asynchronous Stochastic Approximation and Q-Learning, *Machine Learning* **16** (3), pp. 185–202, 1994. DOI: 10.1023/A:102268912 Cited on p. 54, 73, 79

[Van67] J. Vandergraft, Newton's Method for Convex Operators in Partially Ordered Spaces, *SIAM Journal on Numerical Analysis* **4** (3), pp. 406-432, 1967. DOI: 10.1137/0704037 Cited on p.

[Ver18] R. Vershynin, *High-Dimensional Probability, An Introduction with Applications in Data Science*, Cambridge University Press, 2018. ISBN: 9781108231596 Cited on p. 128

[Wai19a] M. Wainwright, *High-Dimensional Statistics, A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019. isbn: 9781108627771 Cited on p. 89

[Wai19b] M. Wainwright, Stochastic approximation with cone-contractive operators: Sharp ℓ_∞-bounds for Q-learning, 2019. arXiv:1905.06265 Cited on p. 69

[Wat89] C. Watkins, *Learning from Delayed Rewards*, PhD Thesis, King's College, Cambridge, UK, 1989. http://www.cs.rhul.ac.uk/~chrisw/new_thesis.pdf Cited on p. 11, 73, 79

Index

S

T

U

V

Credits

The typesetting of this thesis drew inspiration from several different sources. The general thesis format was based on André Miede's Classic Thesis template, as used in Aaron Turon's PhD Thesis. Special thanks to pfasante for sharing his rebuilding attempts on it. The use of flat tables was influenced by Harvey Sheppard's yLaTeX packages and examples. I am grateful to all of them, as well as to the whole TeX Stack Exchange community for already having a solution to virtually any problem that I might had.