Quantitative results for stochastic processes

Thomas Powell
(A report on joint projects with Morenikeji Neri)

University of Bath

Topos Institute Colloquium

16 October 2024

Slides are available on request

Overview

Main topic

Applications of logic (specifically proof theory) in probability and stochastic optimization.

Structure of the talk

1. **Background:** A very high-level overview of applied proof theory ("proof mining").
2. **A simple motivating example:** Monotone convergent sequences.
3. **Main results I:** Quantitative martingale convergence.
4. **Main results II:** Expanding this to general stochastic algorithms.
5. **The future:** This work as part of a much bigger project – *Proof mining in probability*.

A note on syntax/detail

I will present theorems in full detail, but understanding the details is **not important for this talk**! Everything can be understood on a high-level, and I will highlight the important features.

What is applied proof theory?

There is a famous quote due to G. Kreisel (*A Survey of Proof Theory II*):

> *"What more do we know when we know that a theorem can be proved by limited means than if we merely know that it is true?"*

In other words, the **proof** of a theorem gives us much more information than the mere **truth** of that theorem.

Applied proof theory is a branch of logic that uses proof theoretic techniques to exploit this phenomenon.

People do applied proof theory without realising it...

PROBLEM. Give me an upper bound on the nth prime number p_n.

1. What is p_n? I know it exists because of Euclid...
2. Specifically, given $p_1, \ldots, p_{n-1}$, I know that $N := p_1 \cdot \ldots \cdot p_{n-1} + 1$ contains a *new* prime factor q, and so $p_n \leq q \leq N$.
3. In other words, the sequence $\{p_n\}$ satisfies

$$p_n \leq p_1 \cdot \ldots \cdot p_{n-1} + 1 \leq (p_{n-1})^{n-1}$$

4. By induction, it follows that e.g. $p_n < 2^{2^n}$.

This is an extremely simple example of applied proof theory in action! From the **proof** that there are infinitely many primes, we have inferred a **bound** on the nth prime.

… but it's not always that simple

Theorem (Littlewood 1914)

The functions of integers

(a) $\psi(x) - x$, *and*

(b) $\pi(x) - li(x)$

change signs infinitely often, where $\pi(x)$ *is the number of prime* $\leq x$, $\psi(x)$ *is the is logarithm of the l.c.m. of numbers* $\leq x$ *and* $li(x) = \int_0^x (1/\log(u))du$.

The original proof is utterly nonconstructive, using among other things a **case distinction on the Riemann hypothesis**. At the time, no numerical value of x for which $\pi(x) > li(x)$ was known.

In 1952, Kreisel analysed this proof and extracted recursive bounds for sign changes (*On the interpretation of non-finitist proofs, Part II*):

> *"Concerning the bound … note that it is to be expected from our principle, since if the conclusion … holds when the Riemann hypothesis is true, it should also hold when the Riemann hypothesis is nearly true: not all zeros need lie on* $\sigma = \frac{1}{2}$*, but only those whose imaginary part lies below a certain bound … and they need not lie on the line* $\sigma = \frac{1}{2}$*, but near it"*

A very boring example from my own work

Theorem (Kirk and Sims, *Bulletin of the Polish Academy of Sciences* 1999)

Some general, qualitative assumptions: *Suppose that C is a closed subset of a uniformly convex Banach space and $T : C \to C$ is asymptotically nonexpansive with $\mathrm{int}(\mathrm{fix}(T)) \neq \emptyset$. Fix $x \in C$.*

Qualitative conclusion: *The sequence $\{T^n x\}$ converges to a fixed point of T.*

The following is a corollary of a more general quantitative analysis of the above theorem:

Theorem (P., *Journal of Mathematical Analysis and Applications* 2019)

Some concrete, quantitative assumptions: *Let $T : C \to C$ be a nonexpansive mapping in L_p for $2 \leq p < \infty$, and suppose that $B_r[q] \subset \mathrm{fix}(T)$ for some $q \in L_p$ and $r > 0$. Suppose that $x \in C$ and $\|x - q\| < K$.*

A qualitative conclusion: *Define $x_n := T^n x$. Then for any $\varepsilon > 0$ we have*

$$\forall n \geq f(\varepsilon)(\|Tx_n - x_n\| \leq \varepsilon)$$

where

$$f(\varepsilon) := \left\lceil \frac{p \cdot 2^{3p+1} \cdot K^{p+2}}{\varepsilon^p \cdot r^2} \right\rceil$$

Modern applied proof theory

- Origins in the work of Kreisel and the "unwinding" of proofs. Early case studies in number theory.
- Applications in mathematics were brought to maturity by Kohlenbach and his collaborators from late 90s onwards[1].
- There are now hundreds of papers proving new theorems that were obtained using proof theoretic ideas and methods, the majority published in specialised journals in the areas of application, including nonlinear analysis, ergodic theory, convex optimization, . . . (see the proof mining bibliography).
- In parallel, there are logical metatheorems (the first in 2005[2]) that explain individual applications as instances of general logical phenomena.
- Now starting to expand and establish new connections with automated reasoning and formal mathematics.

[1] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics.* **Springer Monographs in Mathematics.** 2008

[2] U. Kohlenbach. *Some logical metatheorems with applications in functional analysis.* **Transactions of the American Mathematical Society**, 2005.

What people working in applied proof theory might do

- Use logical methods to establish quantitative versions of known results in mainstream (non-logic) mathematics.
- Show that a collection of theorems are all instances of a more general, abstract theorem.
- Define new classes of mappings or new types of spaces.
- Develop sophisticated logical systems for reasoning about specific mathematical objects.
- Study a hitherto unexplored area of mathematics to see if proof theoretic methods might be effective and useful. This is **very hard** but **very rewarding** when it works.
- Make important contributions to core logic, including computability theory and theoretical computer science.
- Recently: Formalise their work in a proof assistant (e.g. Lean) or consider automated methods.

Applied proof theory is characterised by thinking about and doing mathematics from a proof-theoretic perspective.

Metastable monotone convergence

Monotone convergence theorem

Theorem

Let $K > 0$ and suppose that $\{x_n\}$ is a monotone sequence of reals with $|x_n| \leq K$ for all $n \in \mathbb{N}$. Then $\{x_n\}$ converges.

Is there a computable, uniform rate of convergence for all sequences in this class? I.e. a function $\phi_K(\varepsilon)$ such that

$$\forall \varepsilon > 0 \, \forall i, j \geq \phi_K(\varepsilon)(|x_i - x_j| < \varepsilon)$$

Specker sequences

There exist monotone bounded sequences of *rational numbers* that do not possess a computable rate of convergence.[3]

We need to consider a different notion of quantitative convergence.

[3] E. Specker. *Nicht konstruktiv beweisbare Sätze der Analysis.* **Journal of Symbolic Logic.** 1949.

A logical approach to the MCT – The statement

The following steps are entirely logic-based (i.e. have nothing to do with convergence):

$$
\begin{aligned}
MCT &:= \forall \varepsilon > 0\, \exists n\, \forall i, j \geq n\, (|x_i - x_j| < \varepsilon)\\
&\iff \forall \varepsilon > 0\, \exists n\, \forall k\, \forall i, j \in [n; n+k]\, (|x_i - x_j| < \varepsilon)\\
&\iff \neg\neg \forall \varepsilon > 0\, \exists n\, \forall k\, \forall i, j \in [n; n+k]\, (|x_i - x_j| < \varepsilon)\\
&\iff \neg \exists \varepsilon > 0\, \forall n\, \exists k\, \exists i, j \in [n; n+k]\, (|x_i - x_j| \geq \varepsilon)\\
&\iff \neg \exists \varepsilon > 0\, \exists g : \mathbb{N} \to \mathbb{N}\, \forall n\, \exists i, j \in [n; n+g(n)]\, (|x_i - x_j| \geq \varepsilon)\\
&\iff \forall \varepsilon > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n\, \forall i, j \in [n; n+g(n)]\, (|x_i - x_j| < \varepsilon) \; := MCT^*
\end{aligned}
$$

Theorem (Metastable monotone convergence theorem – first version)

Let $K > 0$ and suppose that $\{x_n\}$ is a monotone sequence of reals with $|x_n| \leq K$ for all $n \in \mathbb{N}$. Then for any $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$ there exists some $n \in \mathbb{N}$ such that $|x_i - x_j| < \varepsilon$ for all $i, j \in [n; n+g(n)]$.

Question: Can we compute n in ε and g?

A logical approach to the MCT – The proof

Suppose that $\{x_n\} \subset [-K, K]$ is monotone but not Cauchy. Then there is some $\varepsilon > 0$ such that for all $n \in \mathbb{N}$, we can find $k \in \mathbb{N}$ with:

$$\exists i, j \in [n; n+k] \, (|x_i - x_j| \geq \varepsilon)$$

Let $g : \mathbb{N} \to \mathbb{N}$ be a function that finds such a k in terms of n i.e. for all $n \in \mathbb{N}$:

$$\exists i, j \in [n; \tilde{g}(n)] \, (|x_i - x_j| \geq \varepsilon)$$

for $\tilde{g}(x) := x + g(x)$.

Then iterating $\tilde{g}$, for all $e \in \mathbb{N}$:

$$\exists i, j \in [\tilde{g}^{(e)}(0); \tilde{g}^{(e+1)}(0)] \, (|x_i - x_j| \geq \varepsilon) \tag{$*$}$$

In other words, in each of the intervals

$$[0; \tilde{g}(0)], [\tilde{g}(0); \tilde{g}^{(2)}(0)], [\tilde{g}^{(2)}(0); \tilde{g}^{(3)}(0)], \ldots$$

the sequence $\{x_n\}$ experiences a distinct ε-jump (or *fluctuation*).

But a monotone sequence in $[-K, K]$ can experience at most $2K/\varepsilon$ distinct ε-fluctuations, so $(*)$ must fail for some $e \leq \lceil 2K/\varepsilon \rceil$. **Contradiction!**

A logical approach to the MCT – The payoff

Theorem (Metastable monotone convergence theorem)

Take $K, \varepsilon > 0$ *and* $g : \mathbb{N} \to \mathbb{N}$*. Then there exists some* $N \in \mathbb{N}$ *(depending only on* K, ε *and* g*) such that for any monotone sequence* $\{x_n\}$ *in* $[-K, K]$*, there exists* $n \leq N$ *such that* $|x_i - x_j| < \varepsilon$ *for all* $i, j \in [n; n + g(n)]$*. Moreover, we can define*

$$N_K(\varepsilon, g) := \tilde{g}^{(\lceil 2K/\varepsilon \rceil)}(0)$$

for $\tilde{g}(x) := x + g(x)$*.*

Theorem (Generalised metastable convergence theorem)

Take $\phi : (0, 1) \to \mathbb{R}$*,* $\varepsilon > 0$ *and* $g : \mathbb{N} \to \mathbb{N}$*. Then there exists some* $N \in \mathbb{N}$ *such that* ***for any sequence*** $\{x_n\}$ ***in some metric space*** (X, d) ***that experiences at most*** $\phi(\varepsilon)$ ***distinct*** ε***-fluctuations****, there exists* $n \leq N$ *such that* $d(x_i, x_j) < \varepsilon$ *for all* $i, j \in [n; n + g(n)]$*. Moreover, we can define*

$$N_\phi(\varepsilon, g) := \tilde{g}^{(\lceil \phi(\varepsilon) \rceil)}(0)$$

for $\tilde{g}(x) := x + g(x)$*.*

Blog post by Tao[4]

Finite convergence principle. If $\varepsilon > 0$ and $F : \mathbb{Z}_+ \to \mathbb{Z}_+$ is a function and $0 \le x_1 \le x_2 \le \ldots \le x_M \le 1$ is such that M is sufficiently large depending on F and ε, then there exists $1 \le N < N + F(N) \le M$ such that $|x_n - x_m| \le \varepsilon$ for all $N \le n, m \le N + F(N)$.

This principle is easily proven by appealing to the first pigeonhole principle with the sparsified sequence $x_{i_1}, x_{i_2}, x_{i_3}, \ldots$ where the indices are defined recursively by $i_1 := 1$ and $i_{j+1} := i_j + F(i_j)$. This gives an explicit bound on M as $M := i_{\lfloor 1/\varepsilon \rfloor + 1}$. Note that the first pigeonhole principle corresponds to the case $F(N) \equiv 1$, the second pigeonhole principle to the case $F(N) \equiv k$, and the third to the case $F(N) \equiv N$. A particularly useful case for applications is when F grows exponentially in N, in which case M grows tower-exponentially in $1/\varepsilon$.

Informally, the above principle asserts that any sufficiently long (but finite) bounded monotone sequence will experience arbitrarily high-quality amounts of metastability with a specified error tolerance ε, in which the duration F(N) of the metastability exceeds the time N of onset of the metastability by an arbitrary function F which is specified in advance.

Let us now convince ourselves that this is the true finitary version of the infinite convergence principle, by deducing them from each other:

[4] T. Tao. *Soft analysis, hard analysis, and the finite convergence principle.* Essay posted 23 May 2007, appeared in **Structure and Randomness: Pages from Year One of a Mathematical Blog.** 2008.

Paper by Tao[5]

Theorem 1.6 (Finitary norm convergence). *Let $l \geq 1$ be an integer, let $F : \mathbf{N} \to \mathbf{N}$ be a function, and let $\varepsilon > 0$. Then there exists an integer $M^* > 0$ with the following property: If $P \geq 1$ and $f_1, \ldots, f_l : \mathbf{Z}_P^l \to [-1, 1]$ are arbitrary functions on $\mathbf{Z}_P^l$, then there exists an integer $1 \leq M \leq M^*$ such that we have the "L^2 metastability"*

$$\|A_N(f_1, \ldots, f_l) - A_{N'}(f_1, \ldots, f_l)\|_{L^2(\mathbf{Z}_P^l)} \leq \varepsilon \quad (1)$$

for all $M \leq N, N' \leq F(M)$, where we give $\mathbf{Z}_P^l$ the uniform probability measure.

Remark 1.7. For applications, Theorem 1.6 is only of interest in the regime where $F(M)$ is much larger than M, and P is extremely large compared to l, F, or ε. The key points are that the function F is arbitrary (thus one has arbitrarily high quality regions of L^2 metastability), and that the upper bound M^* on M is independent of P. The $l = 1$ version of this theorem was essentially established (with $\mathbf{Z}_P^l$ replaced by an arbitrary measure-preserving system) in [1].

Later there is a footnote...

[1] In proof theory, this finitisation is known as the *Gödel functional interpretation* of the infinitary statement, which is also closely related to the *Kriesel no-counterexample interpretation* [14], [15] or *Herbrand normal form* of such statements; see [13] for further discussion. We thank Ulrich Kohlenbach for pointing out this connection.

[5] T. Tao. *Norm convergence of multiple ergodic averages for commuting transformations.* **Ergodic Theory and Dynamical Systems.** 2008

Metastable convergence: The main points

- There are natural situations where it is impossible to provide computable rates of convergence.
- Where direct rates are not possible, one can often produce either **fluctuation bounds** or **metastable rates** that are both computable and highly uniform.
- Mathematicians outside of logic are interested in fluctuations and metastability (a lot of references are collected in my recent preprint[6]).
- Researchers in applied proof theory have been able to:
  - Extract explicit fluctuation bounds[7] or metastable rates[8] in many different scenarios;
  - Do this in an abstract and general setting;
  - Use them to obtain *concrete* numerical information e.g. direct rates for $||Tx_n - x_n|| \to 0$;
  - Explain why this is possible from a logical point of view[9].

[6] M. Neri and T.Powell. *On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales.* **arXiv:2406.19979.** 2024.

[7] J. Avigad and J. Rute. *Oscillation and the mean ergodic theorem for uniformly convex Banach spaces.* **Ergodic theory and dynamical systems.** 2014.

[8] Lots of examples in the proof mining bibliography.

[9] U. Kohlenbach and P. Safarik. *Fluctuations, effective learnability and metastability in analysis.* **Annals of Pure and Applied Logic.** 2014.

Metastable martingale convergence

Martingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}$ be a filtration. Let $\{X_n\}$ be a sequence of real-valued random variables adapted to $\{\mathcal{F}_n\}$ (i.e. X_n is $\mathcal{F}_n$-measurable) such that $\mathbb{E}[|X_n|] < \infty$ for all $n \in \mathbb{N}$.

We call $\{X_n\}$ a *martingale* if for all $n \in \mathbb{N}$

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n \text{ almost surely.}$$

It is a *submartingale* if $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n$ and a *supermartingale* if $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n$.

Example

Suppose that Alex repeatedly flips a biased coin, winning one euro with probability p and losing one euro with probability $1 - p$ each time. Let X_n be their fortune after n coin flips.

- If $p = 1/2$ then $\{X_n\}$ is a martingale.
- If $p > 1/2$ then $\{X_n\}$ is a submartingale.
- If $p < 1/2$ then $\{X_n\}$ is a supermartingale.

Martingale convergence

Sub- and supermartingales are the stochastic analogue of monotone sequences.

Theorem (Monotone convergence theorem – *first weeks of a first course in analysis*)

Let $K > 0$ and suppose that

- $\{x_n\}$ *is a monotone sequence of reals with*
- $|x_n| \leq K$ *for all* $n \in \mathbb{N}$.

Then $\{x_n\}$ *converges to some real number* x *with* $|x| \leq K$.

Theorem (Doob's L_1-convergence theorem – *usually part of an advanced course on probability and measure*)

Let $K > 0$ and suppose that

- $\{X_n\}$ *is a sub- or supermartingale with*
- $\mathbb{E}[|X_n|] \leq K$ *for all* $n \in \mathbb{N}$.

Then $\{X_n\}$ *converges almost surely to some random variable* X *with* $\mathbb{E}[|X|] \leq K$.

Note: Martingales generalise monotone sequences of reals, so we also cannot expect direct rates of (almost sure) convergence...

A logical approach to Doob - The statement

The following steps are logic-based or use continuity properties of $\mathbb{P}$:

$$
\begin{aligned}
& \mathbb{P}\left(\{X_n\} \text{ converges}\right) = 1 \\
\iff & \mathbb{P}\left(\bigcap_{m=0}^{\infty}\bigcup_{n=0}^{\infty}\bigcap_{k=0}^{\infty} \forall i,j \in [n;n+k]\left(|X_i - X_j| < 2^{-m}\right)\right) = 1 \\
\iff & \forall m \left[\mathbb{P}\left(\bigcup_{n=0}^{\infty}\bigcap_{k=0}^{\infty} \forall i,j \in [n;n+k]\left(|X_i - X_j| < 2^{-m}\right)\right) = 1\right] \\
\iff & \forall m, \lambda > 0\, \exists n \left[\mathbb{P}\left(\bigcap_{k=0}^{\infty} \forall i,j \in [n;n+k]\left(|X_i - X_j| < 2^{-m}\right)\right) > 1 - \lambda\right] \\
\iff & \forall \varepsilon, \lambda > 0\, \exists n\, \forall k\, \left[\mathbb{P}\left(\forall i,j \in [n;n+k]\left(|X_i - X_j| < \varepsilon\right)\right) > 1 - \lambda\right] \\
\iff & \neg\exists \varepsilon, \lambda > 0\, \forall n\, \exists k\, \left[\mathbb{P}\left(\forall i,j \in [n;n+k]\left(|X_i - X_j| < \varepsilon\right)\right) \leq 1 - \lambda\right] \\
\iff & \neg\exists \varepsilon, \lambda > 0\, \exists g : \mathbb{N} \to \mathbb{N}\, \forall n\, \left[\mathbb{P}\left(\forall i,j \in [n;n+g(n)]\left(|X_i - X_j| < \varepsilon\right)\right) \leq 1 - \lambda\right] \\
\iff & \forall \varepsilon, \lambda > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n\, \left[\mathbb{P}\left(\forall i,j \in [n;n+g(n)]\left(|X_i - X_j| < \varepsilon\right)\right) > 1 - \lambda\right]
\end{aligned}
$$

The goal

By analysing the proof of Doob's theorem can we prove the following?

Theorem (Metastable martingale convergence theorem)

Take $K, \varepsilon, \lambda > 0$ and $g : \mathbb{N} \to \mathbb{N}$. Then there exists some $N \in \mathbb{N}$ (depending only on K, ε, λ and g) such that for any sub- or supermartingale $\{X_n\}$ with

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|] < K$$

there exists $n \leq N$ such that

$$|X_i - X_j| < \varepsilon \quad \textit{for all} \quad i, j \in [n; n + g(n)]$$

with probability $> 1 - \lambda$. Moreover, we can define

$$N_K(\lambda, \varepsilon, g) := \ldots$$

Maybe there is also a connection with fluctuations?

For $\varepsilon > 0$ define the random variable $J_\varepsilon(X_n)$ to be the maximum number of ε-fluctuations experienced by the sequence $\{X_n\}$.

Theorem (Neri-P.[10])

For any $\phi : (0,1) \to \mathbb{R}$, $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$ there exists $N \in \mathbb{N}$ such that for any sequence of random variables $\{X_n\}$ with

$$\mathbb{E}\left[J_\varepsilon(X_n)\right] < \phi(\varepsilon)$$

there exists $n \leq N$ such that

$$|X_i - X_j| < \varepsilon \quad \text{for all} \quad i,j \in [n; n+g(n)]$$

with probability $> 1 - \lambda$. Moreover, we can define

$$N_K(\lambda, \varepsilon, g) := \tilde{g}^{(\lceil \phi(\varepsilon)/\lambda \rceil)}(0)$$

for $\tilde{g}(x) := x + g(x)$.

M. Neri and T.Powell. *On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales.* **arXiv:2406.19979.** 2024.

Proof of the theorem (in case there is time!)

Suppose for contradiction that for all $n \in \mathbb{N}$:

$$\mathbb{P}\left(\exists i, j \in [n; n + g(n)](|X_i - X_j| \geq \varepsilon)\right) \geq \lambda \tag{$*$}$$

so in particular, for all $e \in \mathbb{N}$:

$$\mathbb{P}\left(A_e\right) \geq \lambda \quad \text{for} \quad A_e := \exists i, j \in [\tilde{g}^{(e)}(0); \tilde{g}^{(e+1)}(0)](|X_i - X_j| \geq \varepsilon)$$

For any $k \in \mathbb{N}$ we have

$$(k+1)\lambda \leq \sum_{e=0}^{k} \mathbb{P}(A_e) = \sum_{e=0}^{k} \mathbb{E}\left(I_{A_e}\right) = \mathbb{E}\left[\sum_{e=0}^{k} I_{A_e}\right] \leq \mathbb{E}\left[J_\varepsilon(X_n)\right] < \phi(\varepsilon)$$

which is a contradiction for

$$k := \left\lceil \frac{\phi(\varepsilon)}{\lambda} \right\rceil$$

Therefore $\mathbb{P}(A_e) < \lambda$ for some $e \leq k$ and therefore $(*)$ fails for some

$$n \leq \tilde{g}^{(k)}(0)$$

Now it should be easy?

We need a function $\phi_K(\varepsilon)$ such that for any sub- or supermartingale $\{X_n\}$ with

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < K$$

we have

$$\mathbb{E}[J_\varepsilon(X_n)] < \phi(\varepsilon)$$

Theorem (Chashka[11])

For any $K > 0$ there exists a martingale $\{X_n\}$ with

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < K$$

such that

$$\mathbb{E}\left[\sqrt{J_\varepsilon(X_n)}\right] = \infty$$

It turns out you only really get nice fluctuation behaviour for L_2-martingales.

[11] Appears as Theorem 34 of A. G. Kachorovskii. *The rate of convergence in ergodic theorems.* **Russian Mathematical Surveys.** 1996.

For martingales, *crossings* are far easier to characterise

For $a < b$ define the random variable $U_{N,[a,b]}(X_n)$ to be the maximum number of times $\{X_n\}$ upcrosses the interval $[a, b]$ up to time N.

Lemma (Doob's upcrossing inequality for supermartingales)

$$\mathbb{E}\left[U_{\infty,[a,b]}(X_n)\right] \leq \frac{|a| + \mathbb{E}(|X_0|)}{b - a}$$

The inequality encodes the following intuitive idea: *Imagine that $\{X_n\}$ represents a stock, and consider an investment strategy that buys the stock whenever its price falls below a, and sells it whenever its price rises above b. Let Y_N denote your winnings after time N.*

- *Y_N is at least as good as the number of upcrossings $\times\ (b - a)$.*
- *Because $\{X_n\}$ is a supermartingale (i.e. the stock value decreases on average), this strategy can't win on average: $\mathbb{E}[Y_N] \leq 0$.*

Convergence from the upcrossing inequality (very roughly)

If $\{X_n(\omega)\}$ doesn't converge to a limit in $[-\infty, \infty]$, then there exists $a < b$ such that $U_{\infty,[a,b]}(X_n(\omega)) = \infty$, but by the upcrossing inequality $\mathbb{P}\left(\exists a < b\left[U_{\infty,[a,b]}(X_n) = \infty\right]\right) = 0$.

Metastability for L_1-bounded crossings $C_{[a,b]}(X_n)$ (= down + upcrossings)

Theorem (Neri-P.[12])

For any $\lambda, \varepsilon, L, M > 0$ and $g : \mathbb{N} \to \mathbb{N}$ there exists $N \in \mathbb{N}$ such that for any sequence of random variables $\{X_n\}$ such that

$$\mathbb{P}(|X_n| \geq M) < \frac{\lambda}{2} \quad \text{and} \quad \mathbb{E}\left[C_{[a,b]}(X_n)\right] < L \text{ for } [a,b] \in \mathcal{P}(r,l)$$

where $\mathcal{P}(r,l)$ denotes the partition of $[-r,r]$ into l equal subintervals and

$$r := M\left(1 + \frac{2}{p}\right) \quad \text{and} \quad l := p + 2 \quad \text{and} \quad p := \left\lceil \frac{8M}{\varepsilon} \right\rceil$$

there exists $n \leq N$ such that

$$|X_i - X_j| < \varepsilon \quad \text{for all} \quad i, j \in [n; n + g(n)]$$

with probability $> 1 - \lambda$. Moreover, we can define

$$N_{L,M}(\lambda, \varepsilon, g) := \tilde{g}^{(e)}(0) \text{ for } e := \frac{2(p+2)L}{\lambda}$$

M. Neri and T. Powell. *On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales.* **arXiv:2406.19979.** 2024.

A metastable martingale convergence theorem

The following is then a simple corollary:

Theorem (Neri-P.[13])

Take $K, \varepsilon, \lambda > 0$ *and* $g : \mathbb{N} \to \mathbb{N}$. *Then there exists some* $N \in \mathbb{N}$ *(depending only on* K, ε, λ *and* g*) such that for any sub- or supermartingale* $\{X_n\}$ *with*

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|] < K$$

there exists $n \leq N$ *such that*

$$|X_i - X_j| < \varepsilon \quad \textit{for all} \quad i, j \in [n; n + g(n)]$$

with probability $> 1 - \lambda$. *Moreover, we can define*

$$N_K(\lambda, \varepsilon, g) := \tilde{g}^{(e)}(0) \ \textit{for} \ e := c \left(\frac{K}{\lambda \varepsilon} \right)^2$$

where $c > 0$ *is a suitable constant that can be defined explicitly.*

M. Neri and T.Powell. *On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales.* **arXiv:2406.19979.** 2024.

We can use our general framework to do a lot more

Some of our results on martingales:

stochastic process $\{X_n\}$	iterations of $\tilde{g}$
constant, monotone	K/ε
almost sure monotone	$cK/\lambda\varepsilon$
L_2-martingales	$cK^2/\lambda\varepsilon^2$
L_1-martingales	$cK^2/\lambda^2\varepsilon^2$
L_1-almost-martingales	$cK^2/\lambda^{2(1+r)}\varepsilon^2$ some $r \geq 0$

Notes:

- Most of these rates are optimal in a certain sense, but achieving optimal rates and showing that they are optimal was not easy.
- Similar rates can be obtained in other situations where crossing bounds are present e.g. ergodic theory, and we can potentially apply our work to other domains in which crossing inequalities feature[14].

[14] M. Hochman. *Upcrossing inequalities for stationery sequences and applications.* **Annals of Probability.** 2009.

Almost martingales in stochastic optimization

What are "almost supermartingales"?

Theorem (Supermartingale convergence: A simple corollary of Doob's theorem)

Let $\{X_n\}$ be a nonnegative supermartingale i.e.

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n \text{ almost surely.}$$

Then $\{X_n\}$ converges almost surely.

Theorem (Robbins-Siegmund[15], an *almost-supermartingale* convergence theorem)

Let $\{X_n\}$, $\{A_n\}$, $\{B_n\}$ and $\{C_n\}$ be sequences of nonnegative integrable random variables adapted to the filtration $\mathcal{F}_n$ satisfying

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq (1 + A_n)X_n - B_n + C_n \text{ almost surely}$$

where $\sum_{i=0}^{\infty} A_i, \sum_{i=0}^{\infty} C_i < \infty$ almost surely. Then, almost surely, $\{X_n\}$ converges and $\sum_{i=0}^{\infty} B_i < \infty$.

This is one of the most important theorems in stochastic optimization!

[15] Robbins and Siegmund. *A convergence theorem for non negative almost supermartingales and some applications*. **Optimizing methods in statistics**. 1971.

Using the Robbins-Siegmund theorem: A roughly sketched example

Let θ be the unique root of some function M. The Robbins-Monro scheme one of the best known stochastic approximation algorithms, defined by:

$$x_{n+1} = x_n - a_n y_n \quad \text{and} \quad y_n = M(x_n) + \varepsilon_n$$

for $\{\varepsilon_n\}$ some random errors and $\{a_n\}$ some step sizes satisfying $\sum a_n^2 < \infty$ and $\sum a_n = \infty$.

Sketch proof that $x_n \to \theta$ a.s.
For $\mathcal{F}_n := \sigma(x_0, y_0, \ldots, x_n, y_n)$, setting $X_n := (x_n - \theta)^2$ and $V_n := 2M(x_{n+1})(x_{n+1} - \theta)$ we can show[16] that

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq (1 - 2a_{n+1}^2 c^2) X_n - 2a_{n+1} V_n + a_{n+1}^2 (c^2 + K)$$

where $|M(x)| \leq c(|x - \theta| + 1)$ and $\mathbb{E}[\varepsilon_{n+1}^2 \mid \mathcal{F}_n] \leq K$.

In other words, $\{(x_n - \theta)^2\}$ is an almost-supermartingale.

1. By Robbins-Siegmund, $|x_n - \theta|$ converges and $\sum a_n M(x_n)(x_n - \theta) < \infty$ a.s.
2. Since $\sum a_n = \infty$, under suitable conditions on M we can show that $|x_n - \theta| \to 0$.

[16] For details: T. L. Lai. *Stochastic approximation.* **Annals of Statistics.** 2003.

A quantitative Robbins-Siegmund theorem

Theorem (Neri and P.[17])

Let $\{X_n\}$, $\{A_n\}$, $\{B_n\}$ and $\{C_n\}$ be nonnegative integrable stochastic processes adapted to some filtration $\mathcal{F}_n$ such that

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq (1 + A_n)X_n - B_n + C_n$$

almost surely for all $n \in \mathbb{N}$. Suppose that $K > \mathbb{E}[X_0]$ and that $\rho, \tau : (0,1) \to [1, \infty)$ satisfy

$$\mathbb{P}\left(\prod_{i=0}^{\infty}(1 + A_n) \geq \rho(\lambda)\right) < \lambda \ \text{ and } \ \mathbb{P}\left(\sum_{i=0}^{\infty} C_n \geq \sigma(\lambda)\right) < \lambda$$

for all $\lambda \in (0,1)$. Then for any $\varepsilon, \lambda > 0$ and $g : \mathbb{N} \to \mathbb{N}$ there exists some

$$n \leq \tilde{g}^{(e)}(0) \ \text{ for } \ e := c\left(\frac{\rho\left(\frac{\lambda}{8}\right) \cdot \left(K + \sigma\left(\frac{\lambda}{16}\right)\right)}{\lambda\varepsilon}\right)^2 \ \text{ and } \ \tilde{g}(j) := j + g(j)$$

such that

$$|X_i - X_j| < \varepsilon \quad \text{for all} \quad i, j \in [n; n + g(n)]$$

with probability $> 1 - \lambda$. ***(We have an analogous result for $\sum B_n < \infty$).***

[17] Neri and Powell. *A quantitative Robbins-Siegmund theorem.* **Preprint.** 2024.

Questions

Can we use our abstract, quantitative Robbins-Siegmund theorem, and/or other approaches in this spirit, to obtain *useful* numerical information for (classes of) stochastic approximation algorithms?

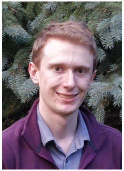
Are there interesting applications in e.g. machine learning?

This is the sort of thing that I and my collaborators are currently working on.

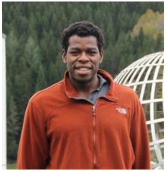
(The answer to both questions is YES!)

The future: Proof mining in probability theory

Proof mining in probability: Collaborators



Ben Langton (Bath)

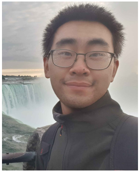


Keji Neri (Bath)

Pedro Pinto (Darmstadt)

Nicholas Pischke (Darmstadt)

Me!



Alex Wan (Bath)

Why is proof mining in probability interesting and rewarding?

- Important theorems are often very simple to state, but have deep and interesting proofs.
- Numerical information (e.g. convergence rates, bounds on constants etc.) can be relevant and sought after by probability theorists (e.g. Berry-Esseen theorem).
- There are many variations of key proof tactics in different settings (e.g. *"reduce to Doob's martingale convergence theorem"*):
  - New quantitative information related to those tactics is then broadly relevant, and
  - proof theoretic insights could also lead to generalisations and unification.
- Probability theory, particulary stochastic convergence, underlies several very active research areas, including *stochastic optimization* and *machine learning*.
- Numerical information is typically *highly uniform*, and there is an exciting prospect of new logical metatheorems that explain this (in fact this is already underway[18]).
- Probability is an extremely beautiful area of mathematics, and it's nice to have an excuse to study it...

[18] Neri and Pischke. *Proof mining and probability theory*. **arXiv:2403.00659** 2024.

Progress so far

Covered in this talk:

- A broad understanding of martingales (and related things) from a computational perspective[19].
- A quantitative Robbins-Siegmund theorem, plus a toolkit for obtaining metastable rates for general almost-supermartingales[20].

Recent work by collaborators:

- A beautiful "proof-theoretically tame" logical system for probability, and a metatheorem that guarantees the extractability of numerical information that is *independent* of the underlying probability space [21].
- New convergence rates for strong laws of large numbers [22].

[19] M. Neri and T.Powell. *On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales.* **arXiv:2406.19979** 2024.

[20] Neri and Powell. *A quantitative Robbins-Siegmund theorem*. **Preprint.** 2024.

[21] Neri and Pischke. *Proof mining and probability theory*. **arXiv:2403.00659** 2024.

[22] Neri. *Quantitative strong laws of large numbers.* **arXiv:2406.19166** 2024.

Ongoing and future work

1. The quantitative study of stochastic algorithms, with applications in stochastic optimization and machine learning.
2. Abstract convergence proofs for generalised classes of algorithms in these areas.
3. Expanding existing logic systems to include an abstract, logical treatment of random variables and notions of integrability.
4. Using convergence results on almost-supermartingales as the basis for a major effort to build a library of computer formalised proofs for stochastic optimization[23].
5. The development of algorithms for automating the reduction to a supermartingale i.e. automatically generating convergence proofs.

and much more ...

THANK YOU!

[23] For some speculative ideas on the formalisation of applied proof theory in general see A. Koutsoukou-Argyraki. *On preserving the computational content of mathematical proofs: Toy examples for a formalising strategy.* **Proceedings of CiE.** 2021.