Simulation methods for lower previsions

Matthias C. M. Troffaes

work partially supported by H2020 Marie Curie ITN, UTOPIAE, Grant Agreement No. 722734

Durham University, United Kingdom

July, 2018

Outline

Outline

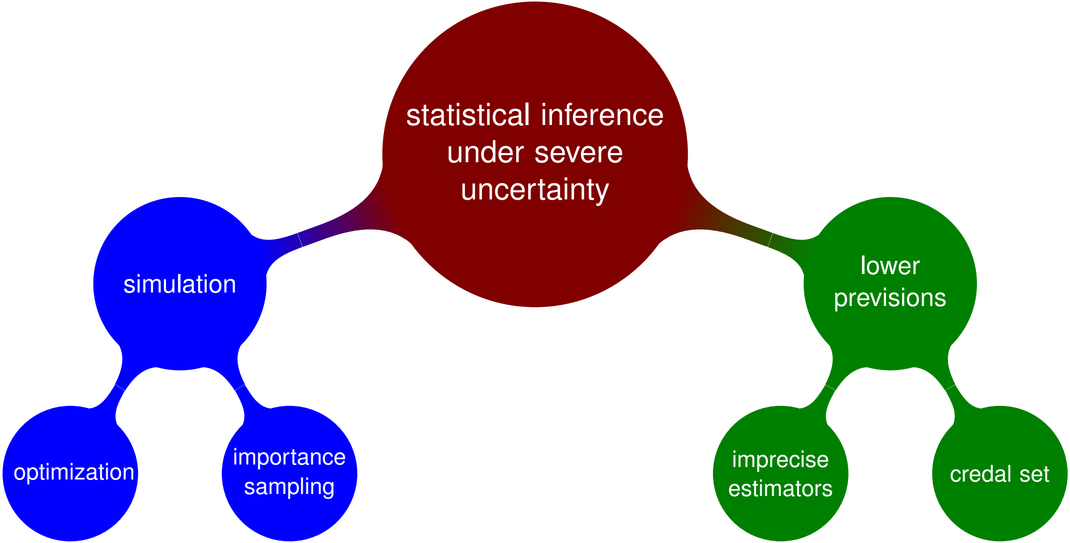
Problem Description

Remember the natural extension of a gamble g:

$$\underline{E}(g) := \min_{p \in \mathcal{M}} E_p(g) \tag{1}$$

- It represents the supremum buying price α you should be willing to pay for g
- We can use this natural extension for all statistical inference and decision making.
- **how to evaluate the minimum in eq. (1) provided we have an estimator for $E_p(g)$?**

Problem Description



Outline

Lower and Upper Estimators for the Minimum of a Function

(see [12])

- Ω = random variable, taking values in some subset of $\mathbb{R}^k$
- t = parameter taking values in some set $\mathcal{T}$ (assume $\mathcal{T}$ countable)
- $\theta(t)$ = arbitrary function of t
- $\hat{\theta}_\Omega(t)$ = arbitrary estimator for θ:

$$E(\hat{\theta}_\Omega(t)) = \theta(t), \tag{2}$$

Aim

Construct an estimator for the minimum of the function θ:

$$\theta_* := \inf_{t \in \mathcal{T}} \theta(t). \tag{3}$$

Example

Say for instance $\mathcal{M} = \{p_t : t \in \mathcal{T}\}$, and let $\theta(t) := E_{p_t}(f)$.
Then $\theta_* = \underline{E}(f)$. So estimation of θ_* = estimation of natural extension.

Lower and Upper Estimators for the Minimum of a Function

Define the function

$$\tau_\Omega \in \arg \inf_{t \in \mathcal{T}} \hat{\theta}_\Omega(t) \tag{4}$$

Theorem (Lower and Upper Estimator Theorem [12])

Assume Ω and Ω' are i.i.d. and let

$$\hat{\theta}_*(\Omega) := \hat{\theta}_\Omega(\tau_\Omega) = \inf_{t \in \mathcal{T}} \hat{\theta}_\Omega(t) \tag{5}$$

$$\hat{\theta}^*(\Omega, \Omega') := \hat{\theta}_\Omega(\tau_{\Omega'}) \tag{6}$$

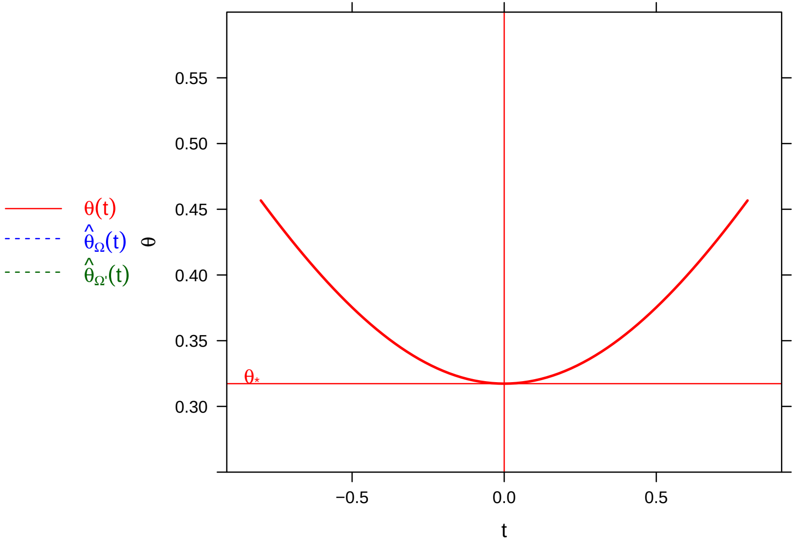
Then

$$\hat{\theta}_*(\Omega) \leq \hat{\theta}^*(\Omega, \Omega') \tag{7}$$

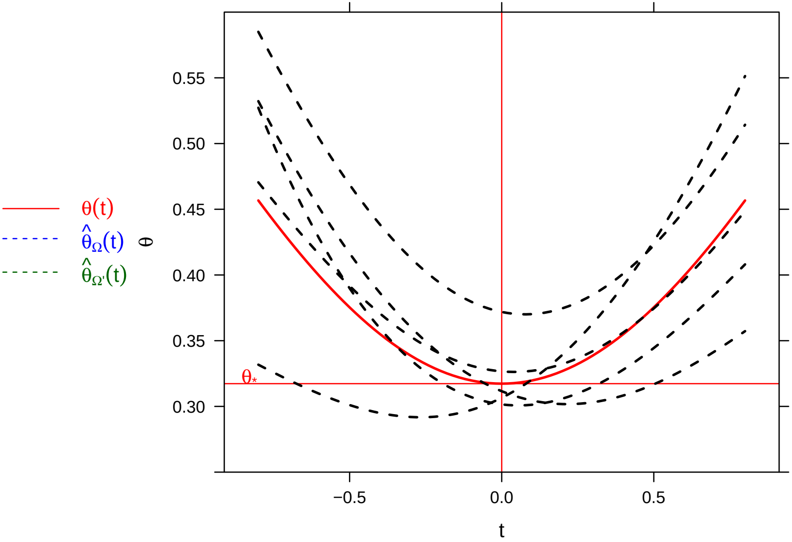
and

$$E(\hat{\theta}_*(\Omega)) \leq \theta_* \leq E(\hat{\theta}^*(\Omega, \Omega')). \tag{8}$$

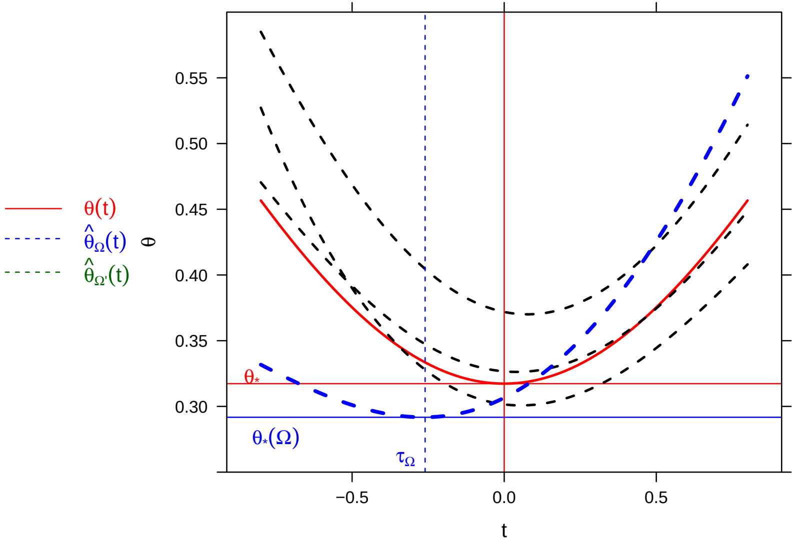
Lower and Upper Estimators for the Minimum of a Function



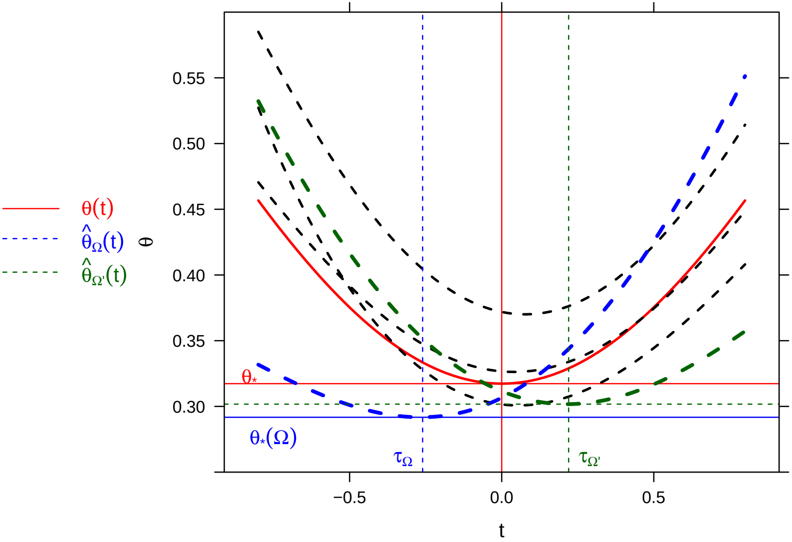
Lower and Upper Estimators for the Minimum of a Function



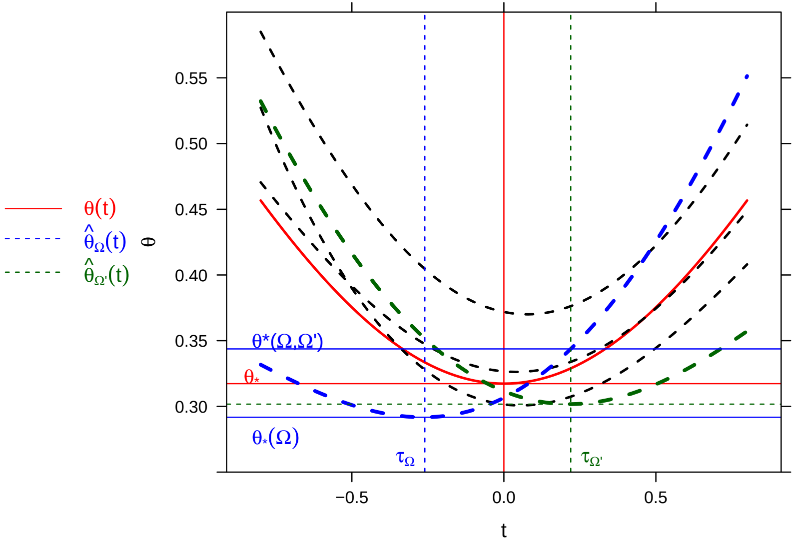
Lower and Upper Estimators for the Minimum of a Function



Lower and Upper Estimators for the Minimum of a Function



Lower and Upper Estimators for the Minimum of a Function



Bias of Lower and Upper Estimators

- $\hat{\theta}_*(\Omega)$: used throughout the literature as an estimator for lower previsions
  not normally noted in the literature that it is negatively biased
  bias can be very large in general (even infinity)!
- $\hat{\theta}^*(\Omega, \Omega')$: introduced at last year's WPMSIIP
  still cannot yet prove much about it
  it allows us to bound the bias without having to do hardcore stochastic process theory

Theorem (Unbiased Case [12])

If there is a $t^* \in \mathcal{T}$ such that $\hat{\theta}_\Omega(t^*) \leq \hat{\theta}_\Omega(t)$ for all $t \in \mathcal{T}$, then

$$\hat{\theta}_*(\Omega) = \hat{\theta}^*(\Omega, \Omega') = \hat{\theta}_\Omega(t^*) \tag{9}$$

and consequently,

$$E(\hat{\theta}_*(\Omega)) = \theta_* = E(\hat{\theta}^*(\Omega, \Omega')). \tag{10}$$

(Condition not normally satisfied, but explains why it is a sensible choice.)

Consistency of the Lower Estimator

Very often, an estimator may take the form of an empirical mean:

$$\hat{\theta}_{\Omega,n}(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{V_i}(t) \tag{11}$$

where $\Omega := (V_i)_{i\in\mathbb{N}}$ and V_i are i.i.d. Under mild conditions, this estimator is consistent:

$$\lim_{n\to\infty} P(|\hat{\theta}_{\Omega,n}(t) - \theta(t)| > \epsilon) = 0 \tag{12}$$

- Under what conditions is $\hat{\theta}_{*n}(\Omega)$ a consistent estimator for θ_*, i.e. when do we have that

$$\lim_{n\to\infty} P(|\hat{\theta}_{*n}(\Omega) - \theta_*| > \epsilon) = 0 \tag{13}$$

- How large should n be?

Consistency of the Lower Estimator

Simple case first:

Theorem (Consistency: Finite Case [12])

If $\mathcal{T}$ is finite, then $\hat{\theta}_{*n}(\Omega)$ is a consistent estimator for θ_*.

(Even though consistent, may require excessively large *n* to control bias!)

General case, no positive answer in general, but consistency can be linked to a well-known condition in stochastic process theory:

Theorem (Consistency: Sufficient Condition for General Case [12])

If the set of functions $\{\hat{\theta}(\cdot, t) : t \in \mathcal{T}\}$ is a Glivenko-Cantelli class, then $\hat{\theta}_{*n}(\Omega)$ is a consistent estimator for θ_*.

Discrepancy Bounds for the Lower Estimator

Notation:

$$Z_n(t) := \hat{\theta}_{\Omega,n}(t) - \theta(t) \tag{14}$$

$$d_n(s,t) := \sqrt{E\left((Z_n(s) - Z_n(t))^2\right)} \tag{15}$$

$$\Delta_n(A) := \sup_{s,t \in A} d_n(s,t) \tag{16}$$

$$\sigma_n^2 := \inf_{t \in \mathcal{T}} \mathrm{Var}(Z_n(t)) = \inf_{t \in \mathcal{T}} \mathrm{Var}(\hat{\theta}_{\Omega,n}(t)) \tag{17}$$

Definition (Talagrand Functional)

Define the Talagrand functional [10, p. 25] as:

$$\gamma_2(\mathcal{T}, d_n) := \inf_{\mathcal{A}_k} \sup_{t \in T} \sum_{k=0}^{\infty} 2^{k/2} \Delta_n(A_k(t)) \tag{18}$$

where the infimum is taken over all 'admissible sequences of partitions of T'.

Discrepancy Bounds for Empirical Mean Lower Estimator

Theorem (Discrepancy Bounds for Empirical Mean Lower Estimator [12])

Assume $\hat{\theta}_{*n}(\Omega) := \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{V_i}(t)$*. There is a universal constant* $L > 0$ *such that, if* $\hat{\theta}_{\Omega,n}(t)$ *is sub-Gaussian, then*

$$P\left(|\hat{\theta}_{*n}(\Omega) - \theta_*| > u(\sigma_1 + \gamma_2(\mathcal{T}, d_1))\right) \leq L\exp(-\tfrac{nu^2}{2}) \tag{19}$$

and

$$E\left(|\hat{\theta}_{*n}(\Omega) - \theta_*|\right) \leq L\frac{\sigma_1 + \gamma_2(\mathcal{T}, d_1)}{\sqrt{n}}. \tag{20}$$

Corollary (Consistency of Empirical Mean Lower Estimator [12])

If $\hat{\theta}_{\Omega,n}(t)$ *is sub-Gaussian, then* $\hat{\theta}_{*n}(\Omega)$ *is a consistent estimator for* θ_* *whenever the minimal standard deviation* σ_1 *and the Talagrand functional* $\gamma_2(\mathcal{T}', d_1)$ *are finite.*

Issue: it is not easy to compute or to bound the Talagrand functional!

Empirical Mean Lower Estimator: How To Achieve Low Bias

Inconsistency Example

- $\hat{\theta}_{\Omega,n}(t)$ has non-zero variance across all t
- $\hat{\theta}_{\Omega,n}(s)$ and $\hat{\theta}_{\Omega,n}(t)$ are independent for all $s \neq t$
- $\mathcal{T}$ is infinite

Then the Talagrand functional $\gamma_2(\mathcal{T}, d_1)$ is $+\infty$.

Important for 2-level Monte Carlo: don't use i.i.d. samples in outer loop over $t \in \mathcal{T}$!

Main Take-Home Message for Design of Estimators

To get a low Talagrand functional (and hence a low bias),
we want $\hat{\theta}_{\Omega,n}(s)$ and $\hat{\theta}_{\Omega,n}(t)$ to be as correlated as possible for all $s \neq t$.

Confidence Interval

Theorem (Confidence Interval from Lower and Upper Estimators [12])

Let $\chi_1, \ldots, \chi_N, \chi'_1, \ldots, \chi'_N$ be a sequence of i.i.d. realisations of Ω. Define

$$Y_* := (\hat{\theta}_*(\chi_i))_{i=1}^N \qquad Y^* := (\hat{\theta}^*(\chi_i, \chi'_i))_{i=1}^N \tag{21}$$

Let $\bar{Y}_*$ and $\bar{Y}^*$ be the sample means of these sequences, and let S_* and S^* be their sample standard deviations. Let t_{N-1} denote the usual two-sided critical value of the t-distribution with $N-1$ degrees of freedom at confidence level $1-\alpha$. Then, provided that $\sup_{x,t} |\hat{\theta}(x,t)| < +\infty$,

$$\left[\bar{Y}_* - t_{N-1}\frac{S_*}{\sqrt{N}}, \bar{Y}^* + t_{N-1}\frac{S^*}{\sqrt{N}}\right] \tag{22}$$

is an approximate confidence interval for θ_* with confidence level (at least) $1-\alpha$.

Why is this rather slow?

Note: we can cheat and use $\hat{\theta}^*(\chi'_i, \chi_i)$ instead for Y^*.

This trick halves computational time (caveat: need $\bar{Y}_* \leq \bar{Y}^*$ with probability $\simeq 1$).

Outline

Example: Toy Problem

(based on [13])

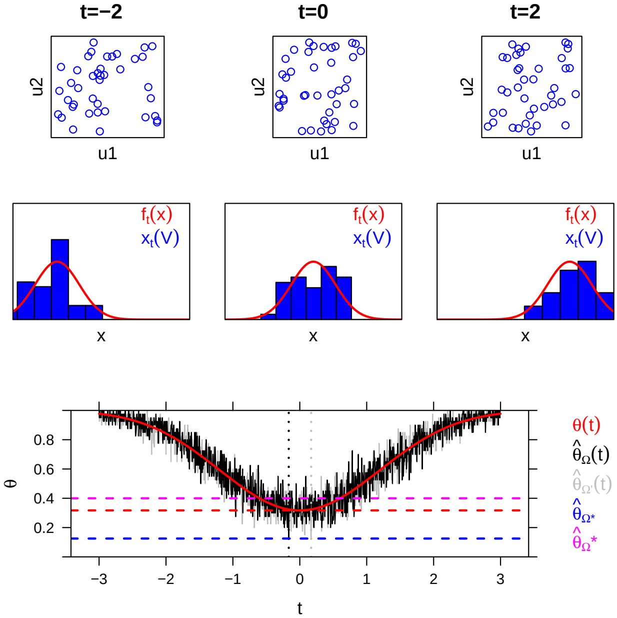
- $V := (U_1, U_2) \sim \mathrm{unif}([0,1]^2)$
- $t := (\mu, \sigma) \in [-3, 3] \times \{1\}$
- $x_t(V) := \mu + \sigma \sqrt{-2 \ln U_1} \cos(2\pi U_2) \sim \mathrm{norm}(\mu, \sigma^2)$
- $f_t(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma}}$
- $h(x) := I_D(x)$ where $D = (-\infty, -1] \cup [1, \infty)$
- $\theta(t) := \int h(x) f_t(x)\, dx$

Example: Two-Level Monte Carlo v1

- different $V_i(t)$ for each value t

$$\hat{\theta}_\Omega(t) := \frac{1}{n}\sum_{i=1}^{n} h(x_t(V_i(t)))$$

- simple
- inefficient
- hard to optimize
- horrible bias
- inconsistent

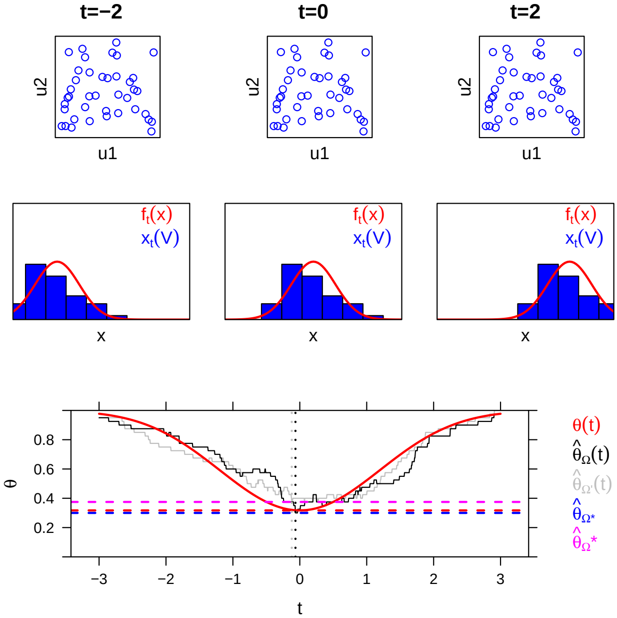


Example: Two-Level Monte Carlo v2

- same V_i for each value t

$$\hat{\theta}_\Omega(t) := \frac{1}{n}\sum_{i=1}^{n} h(x_t(V_i))$$

- most efficient
- can be fairly hard optimize
  might have many local minima
- minimal bias
- consistent



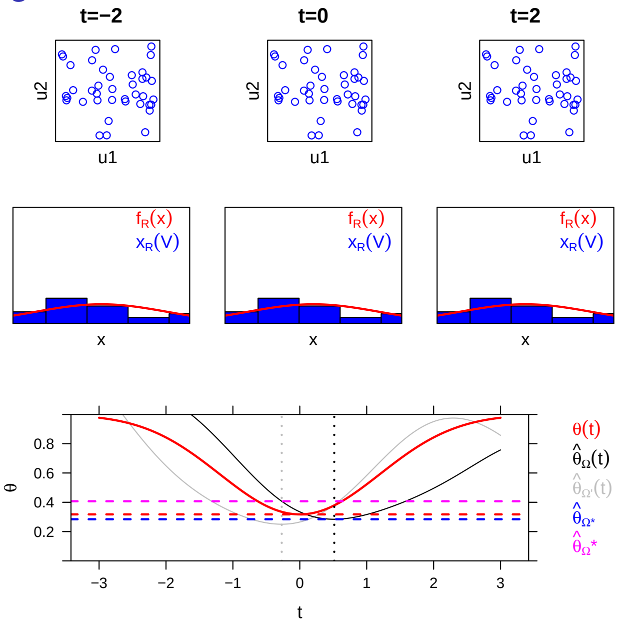
Example: Importance Sampling

(see [8, 4, 14, 11, 3, 12, 13])

- same V_i for each value t
- same samples $x_R(V_i)$ for all t

$$\hat{\theta}_\Omega(t) := \frac{1}{n}\sum_{i=1}^{n} \frac{f_t(x_R(V_i))}{f_R(x_R(V_i))} h(x_R(V_i))$$

- quite efficient for fast densities
- easiest to optimize
- small bias
- still consistent
- f_R needs to cover all f_t
  variance inflation, iterative procedures, ... [13]



Outline

Stochastic Approximation: Kiefer-Wolfowitz

Assume $E(\hat{\theta}_\Omega(t)) = \theta(t)$, uniformly bounded variance. Let

- $a_n := 1/n$
- $c_n := n^{-1/3}$

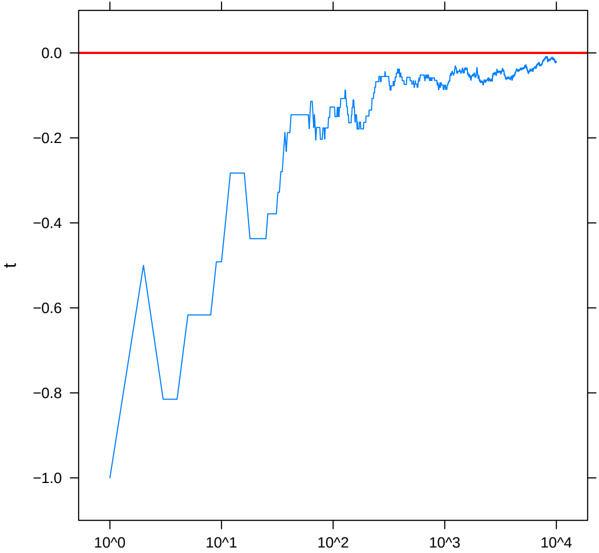
Then

$$t_{n+1}(\Omega_{n+1}) = t_n(\Omega_n) - a_n \underbrace{\frac{\hat{\theta}_{\Omega_{n+1}}(t_n(\Omega_n) + c_n) - \hat{\theta}_{\Omega_{n+1}}(t_n(\Omega_n) - c_n)}{2c_n}}_{\text{stochastic approx of derivative } \frac{d\hat{\theta}}{dt}} \tag{23}$$

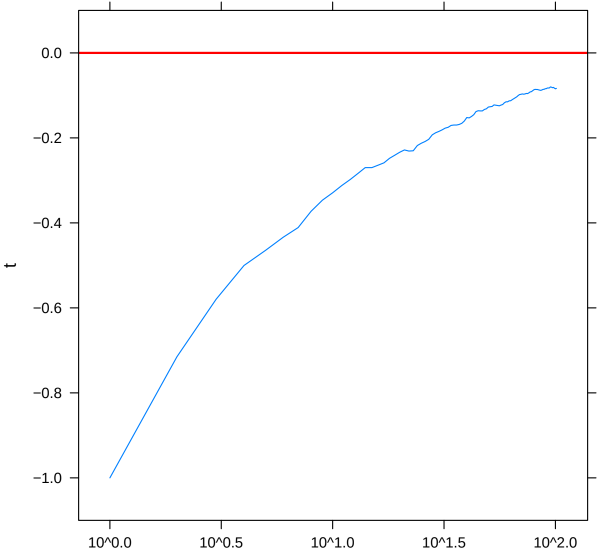
will converge with probability 1 to $\theta_* = \min_t \theta(t)$, provided that $\theta(t)$ is strictly convex.

unbiased and consistent estimator!

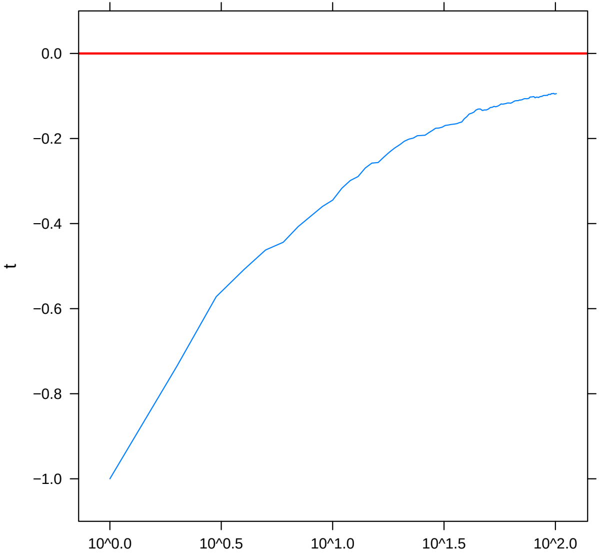
Stochastic Approximation: Example 1 – Single Sample



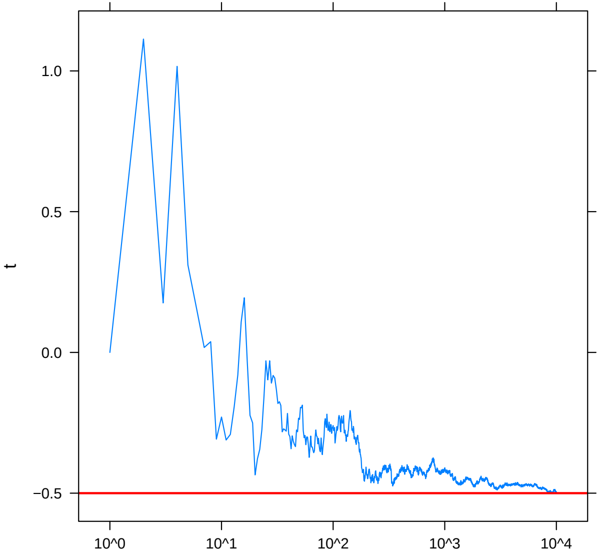
Stochastic Approximation: Example 1 – Mini-Batch MCv2



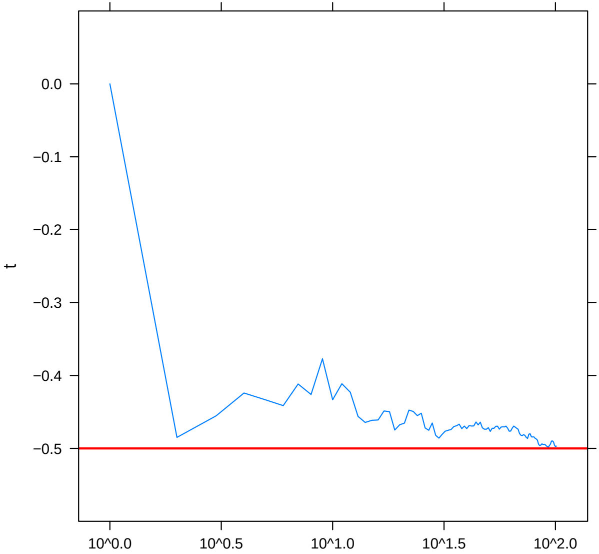
Stochastic Approximation: Example 1 – Mini-Batch Importance



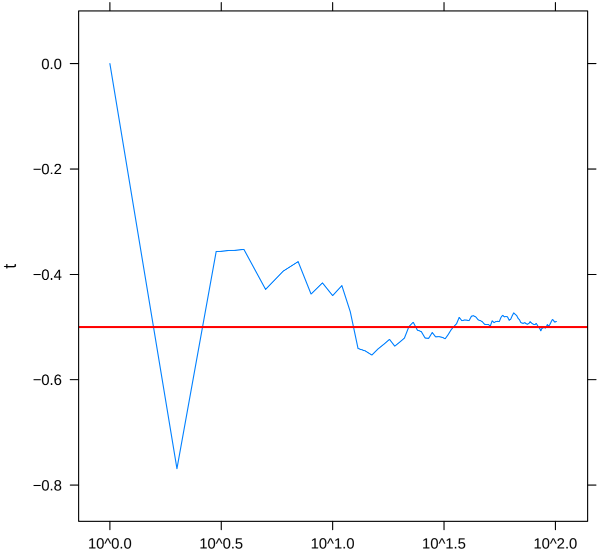
Stochastic Approximation: Example 2 – Single Sample



Stochastic Approximation: Example 2 – Mini-Batch MCv2



Stochastic Approximation: Example 2 – Mini-Batch Importance



Outline

Open Questions

- imprecise estimation
  - the good: we can construct confidence intervals
  - the bad: conditions for consistency hard to quantify
  - the ugly: need multiple runs
- stochastic approximation
  - the good: simple, no bias, consistent
  - the bad: conditions too restrictive? confidence intervals?
  - the ugly: no proofs yet (standard conditions not satisfied yet simulations appear to work)

References I

[1] J. E. Cano, L. D. Hernández, and S. Moral.
Importance sampling algorithms for the propagation of probabilities in belief networks.
International Journal of Approximate Reasoning, 15(1):77–92, 1996.

[2] Marco de Angelis, Edoardo Patelli, and Michael Beer.
Advanced line sampling for efficient robust reliability analysis.
Structural Safety, 52, Part B:170–182, 2015.

[3] Thomas Fetz.
Efficient computation of upper probabilities of failure.
In Christian Bucher, Bruce R. Ellingwood, and Dan M. Frangopol, editors, *12th International Conference on Structural Safety & Reliability*, pages 493–502, 2017.

[4] Thomas Fetz and Michael Oberguggenberger.
Imprecise random variables, random sets, and Monte Carlo simulation.
In Thomas Augustin, Serena Doria, Enrique Miranda, and Erik Quaeghebeur, editors, *ISIPTA '15: Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, pages 137–146, 2015.

[5] Luis D. Hernández and Serafín Moral.
Mixing exact and importance sampling propagation algorithms in dependence graphs.
International Journal of Intelligent Systems, 12(8):553–576, August 1997.

References II

[6] S. Moral and N. Wilson.
Importance sampling algorithms for the calculation of Dempster-Shafer belief.
In *Proceedings of IPMU-96 Conference*, volume 3, pages 1337–1344, 1996.

[7] Michael Oberguggenberger, Julian King, and Bernhard Schmelzer.
Classical and imprecise probability methods for sensitivity analysis in engineering: A case study.
International Journal of Approximate Reasoning, 50(4):680–693, 2009.

[8] B. O'Neill.
Importance sampling for Bayesian sensitivity analysis.
International Journal of Approximate Reasoning, 50(2):270–278, 2009.

[9] Art B. Owen.
Monte Carlo theory, methods and examples.
2013.

[10] Michel Talagrand.
Upper and Lower Bounds for Stochastic Processes.
Springer, 2014.

References III

[11] Matthias C. M. Troffaes.
A note on imprecise Monte Carlo over credal sets via importance sampling.
In Alessandro Antonucci, Giorgio Corani, Inés Couso, and Sébastien Destercke, editors, *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, volume 62 of *Proceedings of Machine Learning Research*, pages 325–332. PMLR, July 2017.

[12] Matthias C. M. Troffaes.
Imprecise Monte Carlo simulation and iterative importance sampling for the estimation of lower previsions.
International Journal of Approximate Reasoning, 101:31–48, October 2018.

[13] Matthias C. M. Troffaes, Thomas Fetz, and Michael Oberguggenberger.
Iterative importance sampling for estimating expectation bounds under partial probability specifications.
In *8th International Workshop on Reliable Engineering Computing: Computing with Confidence (REC 2018)*, pages 139–146, Liverpool, UK, July 2018.

[14] Jiaxin Zhang and Michael D. Shields.
Efficient propagation of imprecise probabilities.
In *7th International Workshop on Reliable Engineering Computing*, pages 197–209, 2016.