A Note on Imprecise Monte Carlo over Credal Sets via Importance Sampling

Matthias C. M. Troffaes

Durham University, United Kingdom

12 July, 2017

Outline

Problem

Importance Sampling

Contributions

- Imprecise Importance Sampling
- Iterative Importance Sampling Method

Example & Simulation Results

Conclusions

Outline

Problem

Notation

(i) set $\mathcal{M} = \{p(\cdot \mid t) \colon t \in \mathcal{T}\}$ of probability density functions

(ii) lower prevision of f:

$$\underline{E}(f) := \min_{t \in \mathcal{T}} \int f(x) p(x \mid t) dx$$

Issues

(i) No closed form for $\int f(x) p(x \mid t) dx$, or expensive to evaluate directly

(ii) $\mathcal{T}$ highly dimensional

Aim

Estimate $\underline{E}(f)$. Key assumptions:

1. Continuous parameterisation: $\mathcal{M} = \{p(\cdot \mid t) \colon t \in \mathcal{T}\}$
2. Can sample from $p(\cdot \mid t)$ for any fixed t
3. Can evaluate $p(x \mid t)$ very fast up to a normalisation constant

Outline

Problem

Importance Sampling

Contributions
- Imprecise Importance Sampling
- Iterative Importance Sampling Method

Example & Simulation Results

Conclusions

Importance Sampling: Basic Ideas

What is Importance Sampling?

Given an i.i.d. sample $x_1, \ldots, x_n \sim p(\cdot \mid \tilde{t})$ for a fixed value of $\tilde{t}$, we can estimate $\int f(x)p(x \mid t)dx$ for all $t \in \mathcal{T}$ simultaneously!

How?

- By reweighting the sample:

$$w'_t(x) = c\frac{p(x \mid t)}{p(x \mid \tilde{t})}$$

- Caveat: the further $p(x \mid t)$ is away from $p(x \mid \tilde{t})$, the worse the estimate!
- Diagnostic: effective sample size

$$n_t := \frac{\left(\sum_{i=1}^n w'_t(x_i)\right)^2}{\sum_{i=1}^n w'_t(x_i)^2}$$

Importance Sampling: Formulas

Self-Normalised Importance Sampling Estimate

$$\int f(x)p(x \mid t)dx \simeq \hat{\mu}_t \pm 1.96\hat{\sigma}_t/\sqrt{n}$$

where

$$\hat{\mu}_t := \frac{\sum_{i=1}^n w_t'(x_i)f(x_i)}{\sum_{i=1}^n w_t'(x_i)} \qquad \hat{\sigma}_t^2 := \frac{1}{n-1}\frac{\frac{1}{n}\sum_{i=1}^n w_t'(x_i)^2(f(x_i)-\hat{\mu}_t)^2}{\left(\frac{1}{n}\sum_{i=1}^n w_t'(x_i)\right)^2}$$

Estimate is a simple non-linear but continuous function of t.

We can optimise $\hat{\mu}_t$ over t!

this is not a new idea: standard non-self-normalised importance sampling already studied by O'Neill, Fetz, Oberguggenberger, Zhang, de Angelis, . . . ; see literature discussion in paper

Outline

Contribution 1: Imprecise Importance Sampling "Does Not Work"

Imprecise Importance Sampling

- Find $t^* := \arg\min_{t\in\mathcal{T}} \hat{\mu}_t$
- Then $\underline{E}(f) \simeq \hat{\mu}_{t^*} \pm 1.96\hat{\sigma}_{t^*} / \sqrt{n}$ provided that $t^* \simeq \arg\min_{t\in\mathcal{T}} E(f \mid t)$

Theoretical Guarantees?

- Normalised case: statistical error (O'Neill), but result is not coherent
- Self-normalised case: result is coherent, but statistical error is open problem

Practical Observations?

- Even in moderately small problems, n_{t^*} is only a very small fraction of n.
- In large models, n_{t^*} is often very close to 1 (i.e. utterly useless).
- Self-normalised imprecise importance sampling
  - is much faster, and
  - is coherent (not true for the non-self-normalised case).
- Sampling distribution does not have to be from $p(x \mid t)$.

Contribution 2: Iterative Importance Sampling Method

Basic Idea

Even though $\hat{\mu}_{t^*}$ can be bad if n_{t^*} is low,
the new t^* is likely still to be an improvement over the original $\tilde{t}$.

Iterative Importance Sampling

(i) Set $\tilde{t}$ to some reasonable initial value in $\mathcal{T}$.

(ii) Generate sample from $p(x \mid \tilde{t})$.

(iii) Find optimal t^* through imprecise importance sampling: $t^* := \arg\min_{t\in\mathcal{T}} \hat{\mu}_t$.

(iv) If $n_{t^*} \simeq n$, stop. Estimate is $\underline{E}(f) \simeq \hat{\mu}_{t^*} \pm 1.96\hat{\sigma}_{t^*} / \sqrt{n}$ (under usual caveat).

(v) If not, set $\tilde{t} = t^*$, and return to item (ii).

Contribution 2: Iterative Importance Sampling Method

Theoretical Guarantees

- Estimate is coherent.
- Convergence? Statistical error? Open problem.

Practical Observations

- Much faster.
- Much lower n required for identical $\hat{\sigma}_{t^*}$.
- Convergences to correct t^* in most (moderately sized) numerical experiments so far.
- Plenty of variations possible (scaling n, scaling $\mathcal{T}$, $\ldots$).

Outline

Example & Simulation Results

Inputs

- Δ = k-dimensional unit simplex, $k = 5$
- $p(x \mid t) := \frac{\Gamma(s)}{\prod_{j=1}^{k} \Gamma(st_j)} \prod_{j=1}^{k} x_j^{st_j - 1}$ with $s = 2$ and $t \in \Delta$
- $w'_t(x) = \prod_{j=1}^{k} x_j^{2(t_j - \tilde{t}_j)} \propto p(x \mid t)/p(x \mid \tilde{t})$
- $\mathcal{T} := \{t \in \Delta : t_j \geq 0.1\}$
- $f(x) = x_1 + 2x_2 + 5x_3 + 4x_4 - 3x_5$ (note, analytically, $E(f) = -0.6$)

Example & Simulation Results

Imprecise Importance Sampling

n	5	50	500	5000
$\hat{\mu}_{t^*}$	1.50	0.13	-0.85	-0.29
$\hat{\sigma}_{t^*}$	0.11	3.18	10.83	10.74
$\hat{\sigma}_{t^*} / \sqrt{n}$	0.048	0.45	0.48	0.15
n_{t^*}	1.104	15.016	6.061	141.67
t_1^*	0.1	0.1	0.17	0.1
t_2^*	0.57	0.1	0.1	0.1
t_3^*	0.1	0.1	0.1	0.1
t_4^*	0.1	0.1	0.1	0.1
t_5^*	0.13	0.6	0.53	0.6

Observations

- For $n = 5000$, simulation takes about 200 seconds.
- Very low n_{t^*}. The $n = 500$ case is particularly dreadful.
- Estimate generally outside confidence interval esp. when n_{t^*} is low.
- In all cases, $\hat{\sigma}_{t^*}$ is an extremely poor estimate of the actual standard deviation.

Example & Simulation Results

Iterative Importance Sampling With $n = 141$

iteration	1	2	3
$\hat{\mu}_{t^*}$	0.062	-0.39	-0.63
$\hat{\sigma}_{t^*}$	4.28	2.00	1.76
$\hat{\sigma}_{t^*} / \sqrt{n}$	0.36	0.17	0.15
n_{t^*}	21.60	105.93	141.00
t_1^*	0.16	0.1	0.1
t_2^*	0.1	0.1	0.1
t_3^*	0.1	0.1	0.1
t_4^*	0.1	0.1	0.1
t_5^*	0.54	0.6	0.6

Observations

- Total simulation takes about 6 seconds (non-self-normalised version: 86 seconds).
- Iteration 2: correct t^* identified; iteration 3: $n_{t^*} = n$, optimisation immediate.
- Final estimate comfortably within confidence interval.
- Accurate estimate also for $\hat{\sigma}_{t^*}$ (analytical value is 1.792577).

Outline

Problem

Importance Sampling

Contributions

- Imprecise Importance Sampling
- Iterative Importance Sampling Method

Example & Simulation Results

Conclusions

Conclusions

Main Conclusions

- importance sampling allows us to estimate lower expectations around an entire neighbourhood of distributions
- self-normalised importance sampling: faster, required for coherence, but theoretically harder to work with; not much studied in imprecise probability setting
- naive imprecise importance sampling severely limited
- novel iterative importance sampling method extremely promising

Enticed? Come and speak to me in the breaks over coffee/lunch!!

Thank you!

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 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.

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

[5] 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.

[6] 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.

References II

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

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

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