Monte Carlo Estimation for Imprecise Probabilities Basic Properties

Arne Decadt Gert de Cooman Jasper De Bock

Monte Carlo Estimation for Imprecise Probabilities Basic Properties

Monte Carlo Estimation for Imprecise Probabilities Basic Properties

Arne Decadt Gert de Cooman Jasper De Bock

Good afternoon everyone. My name is Arne and my promotors are Gert and Jasper. And I want to tell you about how you can use Monte Carlo simulation in the context of imprecise probabilities.

Setting

Monte Carlo

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k^P) \approx \mathrm{E}^P(f)$$

Imprecise Probability

$$\mathscr{P} = \{P_t \colon t \in T\}$$

$$\underline{\mathrm{E}}^{\mathscr{P}}(f) = \inf_t \mathrm{E}^{P_t}(f)$$

Setting
Monte Carlo
$\frac{1}{n}\sum_{k=1}^{n} f(X_k^P) \approx \mathrm{E}^P(f)$
Imprecise Probability
$\mathscr{P} = \{P_t \colon t \in T\}$
$\underline{\mathrm{E}}^{\mathscr{P}}(f) = \inf_t \mathrm{E}^{P_t}(f)$

Like the title says, we combine Classical Monte Carlo and Imprecise probability For a Monte Carlo part, we want to proceed analogous and take a sequence of samples, compute their function values and then take the average. For the Imprecise probability part, well instead of estimating for a single probability measure, we look at a set of probability measures which we will parameterise by parameter t, and then try to calculate the corresponding lower expectation.

Estimators for lower expectations

- $\mathrm{E}^{P_1}(f)$
- $\mathrm{E}^{P_2}(f)$

Estimators for lower expectations

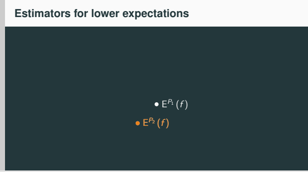
- $\mathrm{E}^{P_1}(f)$
- $\mathrm{E}^{P_2}(f)$

Now if there are two probability measures, it is possible to do two independent Monte Carlo simulations and pick the lowest one and almost trivially – by the law of large numbers – this will converge to the right value.

Estimators for lower expectations

- $E^{P_1}(f)$
- $E^{P_2}(f)$

Monte Carlo Estimation for Imprecise Probabilities Basic Properties



Now if there are two probability measures, it is possible to do two independent Monte Carlo simulations and pick the lowest one and almost trivially – by the law of large numbers – this will converge to the right value.

Estimators for lower expectations

Infinite set of probability measures

So the problem arises when you have an infinite set of probability measures. If you try the same independent sampling, you encounter many problems. First of all, you can't do Monte Carlo for all points, so you would have to define a grid. Secondly, if you have a lot of simulations, the probability will be higher that one of them is an outlier that skews your estimate and thirdly, if you try get a higher accuracy, you need to make your grid finer, but for each new grid point you have to start over. So that is not what they do in practice.

Estimators for lower expectations

1. Fix sampling P

Estimators for lower expectations

1. Fix sampling P

In practice, they will fix a sampling probability measure, not necessarily in our set of probability measures. And then for every parameter t we find functions f_t that have this property. Probably the most popular way to do it – if the probability measures have densities – is importance sampling. This is popular, because often the function f is hard to compute and this technique makes that the function evaluation only has to be calculated once, as Thomas Fetz also discussed yesterday.

There are other ways to choose these functions when the probability measures do not have densities.

But now the question is again, can we approximate the lower expectation by taking the infimum of the expectations? Well we have looked at the properties

Estimators for lower expectations

1. Fix sampling P
2. Find f_t such that $\mathrm{E}^{P}(f_t) = \mathrm{E}^{P_t}(f)$

Estimators for lower expectations

1. Fix sampling P
2. Find f_t such that $\mathrm{E}^{P}(f_t) = \mathrm{E}^{P_t}(f)$

In practice, they will fix a sampling probability measure, not necessarily in our set of probability measures. And then for every parameter t we find functions f_t that have this property. Probably the most popular way to do it – if the probability measures have densities – is importance sampling. This is popular, because often the function f is hard to compute and this technique makes that the function evaluation only has to be calculated once, as Thomas Fetz also discussed yesterday.

There are other ways to choose these functions when the probability measures do not have densities.

But now the question is again, can we approximate the lower expectation by taking the infimum of the expectations? Well we have looked at the properties

Estimators for lower expectations

1. Fix sampling P
2. Find f_t such that $\mathrm{E}^{P}(f_t) = \mathrm{E}^{P_t}(f)$

   $\rightarrow$ Importance Sampling

Estimators for lower expectations
1. Fix sampling P
2. Find f_t such that $\mathrm{E}^{P}(f_t) = \mathrm{E}^{P_t}(f)$
$\rightarrow$ Importance Sampling

In practice, they will fix a sampling probability measure, not necessarily in our set of probability measures. And then for every parameter t we find functions f_t that have this property. Probably the most popular way to do it – if the probability measures have densities – is importance sampling. This is popular, because often the function f is hard to compute and this technique makes that the function evaluation only has to be calculated once, as Thomas Fetz also discussed yesterday.

There are other ways to choose these functions when the probability measures do not have densities.

But now the question is again, can we approximate the lower expectation by taking the infimum of the expectations? Well we have looked at the properties

Estimators for lower expectations

1. Fix sampling P
2. Find f_t such that $\mathrm{E}^P(f_t) = \mathrm{E}^{P_t}(f)$

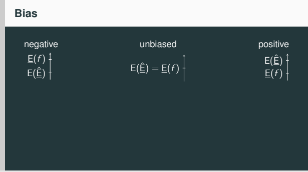
   $\rightarrow$ Importance Sampling

$$\underline{\mathrm{E}}^{\mathscr{P}}(f) \overset{?}{\approx} \inf_t \sum_{k=1}^{n} f_t\left(X_k^P\right) = \underline{\hat{\mathrm{E}}}$$

Estimators for lower expectations
1. Fix sampling P
2. Find f_t such that $\mathrm{E}^P(f_t) = \mathrm{E}^{P_t}(f)$
$\rightarrow$ Importance Sampling
$\underline{\mathrm{E}}^{\mathscr{P}}(f) \overset{?}{\approx} \inf_t \sum_{k=1}^{n} f_t\left(X_k^P\right) = \underline{\hat{\mathrm{E}}}$

In practice, they will fix a sampling probability measure, not necessarily in our set of probability measures. And then for every parameter t we find functions f_t that have this property. Probably the most popular way to do it – if the probability measures have densities – is importance sampling. This is popular, because often the function f is hard to compute and this technique makes that the function evaluation only has to be calculated once, as Thomas Fetz also discussed yesterday.
There are other ways to choose these functions when the probability measures do not have densities.

But now the question is again, can we approximate the lower expectation by taking the infimum of the expectations? Well we have looked at the properties



Bias

negative

$\underline{E}(f)$

$E(\hat{\underline{E}})$

unbiased

$E(\hat{\underline{E}}) = \underline{E}(f)$

positive

$E(\hat{\underline{E}})$

$\underline{E}(f)$

So what is the bias of this lower expectation estimator? Well it underestimates the real value. This is a good thing, as the estimate will be conservative. Now how the bias vary with the sample size? Well it appears that can only get closer to the lower expectation. So it is bounded above and increasing; that sounds like a theorem from calculus; so it has a limit. But is it the right one? Is the estimator asymptotically unbiased? The answer is "not necessarily". And we investigated a stronger notion than asymptotically unbiased, because what we prefer in practice is consistency. It's not the expectation that should go to the right value, we want that our own simulation goes to the right value.

Bias

negative

$\underline{E}(f)$

$E(\hat{\underline{E}})$

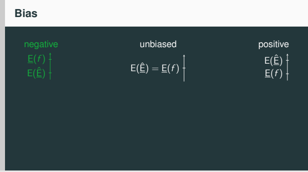
unbiased

$E(\hat{\underline{E}}) = \underline{E}(f)$

positive

$E(\hat{\underline{E}})$

$\underline{E}(f)$



So what is the bias of this lower expectation estimator? Well it underestimates the real value. This is a good thing, as the estimate will be conservative. Now how the bias vary with the sample size? Well it appears that can only get closer to the lower expectation. So it is bounded above and increasing; that sounds like a theorem from calculus; so it has a limit. But is it the right one? Is the estimator asymptotically unbiased? The answer is "not necessarily". And we investigated a stronger notion than asymptotically unbiased, because what we prefer in practice is consistency. It’s not the expectation that should go to the right value, we want that our own simulation goes to the right value.

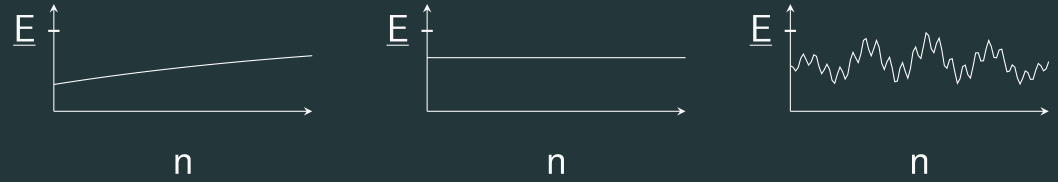
Bias

negative: $\underline{E}(f)$ above $E(\hat{\underline{E}})$

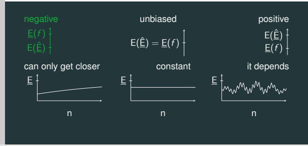
unbiased: $E(\hat{\underline{E}}) = \underline{E}(f)$

positive: $E(\hat{\underline{E}})$ above $\underline{E}(f)$

can only get closer — constant — it depends



Monte Carlo Estimation for Imprecise Probabilities Basic Properties



So what is the bias of this lower expectation estimator? Well it underestimates the real value. This is a good thing, as the estimate will be conservative. Now how the bias vary with the sample size? Well it appears that can only get closer to the lower expectation. So it is bounded above and increasing; that sounds like a theorem from calculus; so it has a limit. But is it the right one? Is the estimator asymptotically unbiased? The answer is "not necessarily". And we investigated a stronger notion than asymptotically unbiased, because what we prefer in practice is consistency. It's not the expectation that should go to the right value, we want that our own simulation goes to the right value.

Bias

negative

$\underline{\mathrm{E}}(f)$

$\mathrm{E}(\hat{\underline{\mathrm{E}}})$

unbiased

$\mathrm{E}(\hat{\underline{\mathrm{E}}}) = \underline{\mathrm{E}}(f)$

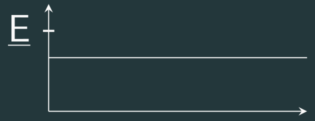
positive

$\mathrm{E}(\hat{\underline{\mathrm{E}}})$

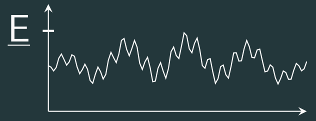
$\underline{\mathrm{E}}(f)$

can only get closer

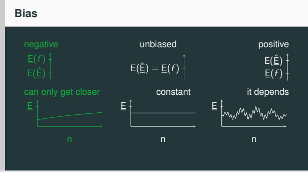
constant



it depends



Monte Carlo Estimation for Imprecise Probabilities Basic Properties



So what is the bias of this lower expectation estimator? Well it underestimates the real value. This is a good thing, as the estimate will be conservative. Now how the bias vary with the sample size? Well it appears that can only get closer to the lower expectation. So it is bounded above and increasing; that sounds like a theorem from calculus; so it has a limit. But is it the right one? Is the estimator asymptotically unbiased? The answer is "not necessarily". And we investigated a stronger notion than asymptotically unbiased, because what we prefer in practice is consistency. It's not the expectation that should go to the right value, we want that our own simulation goes to the right value.

consistency

When you say consistency, you should specify which consistency, convergence in probability or convergence almost surely. The second one is stronger so we chose this one. The first one should not be forgotten though, as it is often easier and is practical to estimate the rate of convergence.

consistency

1. In probability

$$\lim_{n\to\infty} P^{\infty}\left(\left|\hat{\underline{\mathrm{E}}}_n - \mathrm{E}(f)\right| > \epsilon\right) = 0$$

2. Almost surely

$$P^{\infty}\left(\lim_{n\to\infty} \hat{\underline{\mathrm{E}}}_n = \mathrm{E}(f)\right) = 0$$

Monte Carlo Estimation for Imprecise Probabilities Basic Properties

consistency
1. In probability
$$\lim_{n\to\infty} P^{\infty}\left(\left|\hat{\underline{\mathrm{E}}}_n - \mathrm{E}(f)\right| > \epsilon\right) = 0$$
2. Almost surely
$$P^{\infty}\left(\lim_{n\to\infty} \hat{\underline{\mathrm{E}}}_n = \mathrm{E}(f)\right) = 0$$

When you say consistency, you should specify which consistency, convergence in probability or convergence almost surely. The second one is stronger so we chose this one. The first one should not be forgotten though, as it is often easier and is practical to estimate the rate of convergence.

consistency

1. In probability

$$\lim_{n\to\infty} P^\infty\left(\left|\hat{\underline{\mathrm{E}}}_n - \mathrm{E}(f)\right| > \epsilon\right) = 0$$

2. Almost surely

$$P^\infty\left(\lim_{n\to\infty} \hat{\underline{\mathrm{E}}}_n = \mathrm{E}(f)\right) = 0$$

consistency

1. In probability

$$\lim_{n\to\infty} P^\infty\left(\left|\hat{\underline{\mathrm{E}}}_n - \mathrm{E}(f)\right| > \epsilon\right) = 0$$

2. Almost surely

$$P^\infty\left(\lim_{n\to\infty} \hat{\underline{\mathrm{E}}}_n = \mathrm{E}(f)\right) = 0$$

When you say consistency, you should specify which consistency, convergence in probability or convergence almost surely. The second one is stronger so we chose this one. The first one should not be forgotten though, as it is often easier and is practical to estimate the rate of convergence.

Consistency

$$\inf_t \sum_{k=0}^{n} f_t\left(X_k^{P}\right) \xrightarrow{?} \inf_t \mathrm{E}^{P_t}(f) = \underline{\mathrm{E}}^{\mathscr{P}}(f) \qquad \text{as } n \to \infty$$

2019-07-04

Monte Carlo Estimation for Imprecise Probabilities Basic Properties

Consistency

$$\inf_t \sum_{k=0}^{n} f_t\left(X_k^{P}\right) \xrightarrow{?} \inf_t \mathrm{E}^{P_t}(f) = \underline{\mathrm{E}}^{\mathscr{P}}(f) \qquad \text{as } n \to \infty$$

Consistency

$$\inf_t \sum_{k=0}^{n} f_t\left(X_k^P\right) \xrightarrow{?} \inf_t \mathrm{E}^{P_t}(f) = \underline{\mathrm{E}}^{\mathscr{P}}(f) \qquad \text{as } n \to \infty$$

$$\Uparrow$$

$$\sup_t \left| \sum_{k=0}^{n} f_t\left(X_k^P\right) - \mathrm{E}^{P_t}(f) \right| \xrightarrow{?} 0 \qquad \text{as } n \to \infty$$

Consistency

$$\inf_t \sum_{k=0}^{n} f_t\left(X_k^P\right) \xrightarrow{?} \inf_t \mathrm{E}^{P_t}(f) = \underline{\mathrm{E}}^{\mathscr{P}}(f) \qquad \text{as } n \to \infty$$

$$\Uparrow$$

$$\sup_t \left| \sum_{k=0}^{n} f_t\left(X_k^P\right) - \mathrm{E}^{P_t}(f) \right| \xrightarrow{?} 0 \qquad \text{as } n \to \infty$$

Consistency

When is this the case?

Consistency

When is this the case?

Consistency

When is this the case?

- restrictions on size of T

Consistency

When is this the case?

- restrictions on size of T

Consistency

When is this the case?

- restrictions on size of T
- continuity conditions for f_t

2019-07-04

Monte Carlo Estimation for Imprecise Probabilities Basic Properties

Consistency

When is this the case?

- restrictions on size of T
- continuity conditions for f_t

Consistency

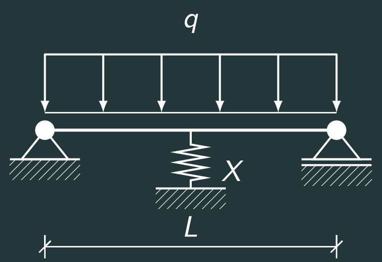
When is this the case?

- restrictions on size of T
- continuity conditions for f_t

finite T	$\mathbb{R}^n \supset T$ compact $p_t(x)$ cont. diff. in (x, t) $\mathrm{E}^P(\sup_{t \in T} p_t) < +\infty$	$\mathbb{R}^n \supset T$ bounded $\lVert \nabla_t p_t(x) \rVert < F(x)$	for all $\epsilon > 0 : T$ has a finite ϵ-cover $\lvert p_t(x) - p_s(x) \rvert \leqslant d(s, t) F(x)$

← easier but more restrictive ——— more general but more complex →

Practical Example

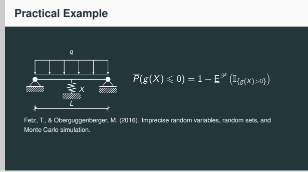


$$\overline{P}(g(X) \leqslant 0) = 1 - \underline{\mathrm{E}}^{\mathscr{P}}\left(\mathbb{I}_{\{g(X)>0\}}\right)$$

Fetz, T., & Oberguggenberger, M. (2016). Imprecise random variables, random sets, and Monte Carlo simulation.

2019-07-04

Monte Carlo Estimation for Imprecise Probabilities Basic Properties



See you at my poster.

See you at my poster.