# Monte Carlo Estimation for Imprecise Probabilities: Basic Properties

**Arne Decadt, Gert de Cooman, Jasper De Bock**
{Arne.Decadt,Gert.deCooman,Jasper.DeBock}@UGent.be

GHENT UNIVERSITY

**FLip, Belgium**

## Problem Statement

We estimate lower expectations $\underline{\mathrm{E}}^{\mathscr{P}}(f) = \inf_{t \in T} \mathrm{E}^{P_t}(f)$ for some set of probability measures $\mathscr{P} = \{P_t : t \in T\}$ and do this using the technique of Monte Carlo. In the classical case, Monte Carlo says that for large sample sizes $n$ we have $\mathrm{E}^P(f) \approx \sum_{k=1}^n f(X_k^P) = \hat{\mathrm{E}}^P(f)$. We consider imprecise extensions of such estimators. In this poster we will use an example where $f(x) = \cos(2x)$, $T = [-2,2]$, $P_t \sim N(t,1)$. For this example it can be calculated exactly that $\underline{\mathrm{E}}^{\mathscr{P}}(f) = -\mathrm{e}^{-2} = \mathrm{E}^{P_{\frac{\pi}{2}}}(f) = \mathrm{E}^{P_{-\frac{\pi}{2}}}(f)$.

## Naive Method

A straightforward attempt at an estimator is to take a representative finite subset $T'$ of $T$ and do independent classical Monte Carlo simulations for every one of them and take their minimum.

$$\hat{\underline{\mathrm{E}}}^{\mathscr{P}}(f) = \min_{t \in T'} \hat{\mathrm{E}}^{P_t}(f)$$

Let $m$ be the number of elements in $T'$. For $m > 1$, we can choose $T'$ equidistant from $-2$ to $2$.

## Transform Method

We consider estimators of the form

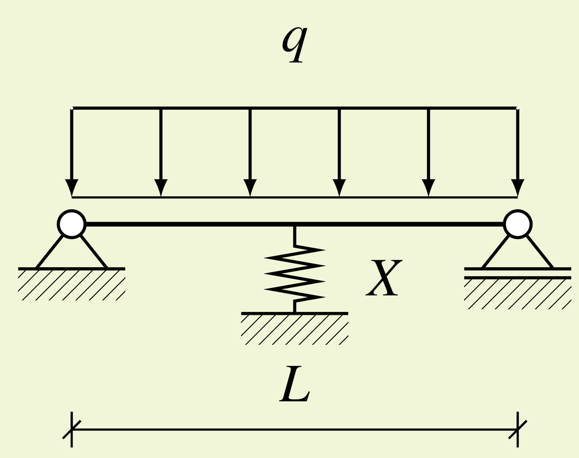
$$\hat{\underline{\mathrm{E}}}_n^{\mathscr{P}}(f) = \inf_{t \in T} \frac{1}{n} \sum_{k=1}^{n} f_t(X_k^P),$$

for chosen functions $f_t$ for which $\mathrm{E}^P(f_t) = \mathrm{E}^{P_t}(f)$.

We give two examples of such functions:

1. Inverse transform sampling: $P$ is the uniform distribution on $(0,1)$. Consider the quantile function $F_{P_t}^{\dagger}$ (the pseudo-inverse of the cdf). Now if $f_t = f \circ F_{P_t}^{\dagger}$ on $(0,1)$, then we have the desired property.
2. Importance sampling: Suppose the probability measures $P$ and $P_t$ have densities $p$ and $p_t$ respectively for every $t \in T$, and for every $t \in T$: $\operatorname{supp} p \supset \operatorname{supp} p_t$. If $f_t = f \cdot \frac{p_t}{p}$ on $\operatorname{supp} p$, then we have the desired property.

## Example from Literature




Thomas Fetz and Michael Oberguggenberger used importance sampling to estimate the upper failure probability of a beam on a spring of unknown stiffness $X$. They assume $X$ distributed normally with with mean and standard deviation $(\mu, \sigma)$ in $[\underline{\mu}, \overline{\mu}] \times [\underline{\sigma}, \overline{\sigma}]$.

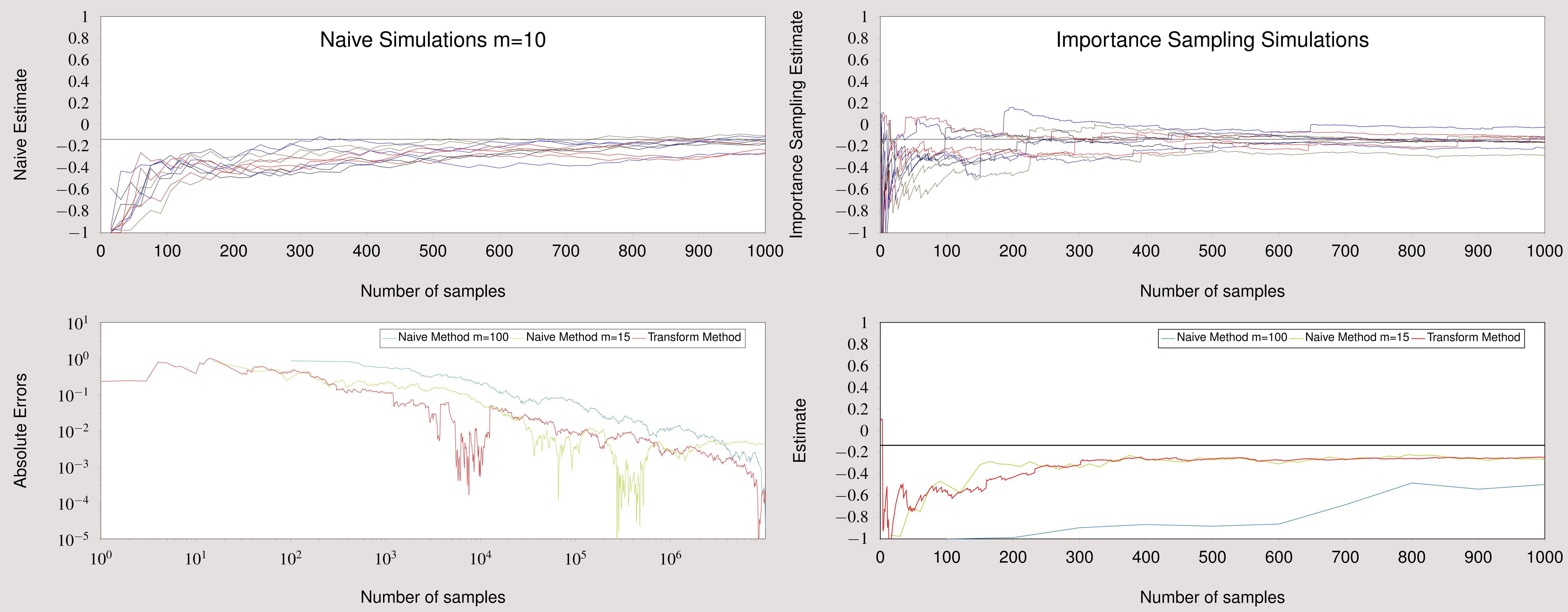
The objective can be rephrased in our context of lower expectations as

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

**Our method proves that their estimator is consistent.**

## Example Simulations

For the importance sampling $P \sim N(0,1)$.




## Bias

1. The bias is negative and the absolute bias decreases with $n$:
$$\mathrm{E}\Big(\hat{\underline{\mathrm{E}}}_{n-1}^{\mathscr{P}}(f)\Big) \leqslant \mathrm{E}\Big(\hat{\underline{\mathrm{E}}}_{n}^{\mathscr{P}}(f)\Big) \leqslant \underline{\mathrm{E}}^{\mathscr{P}}(f).$$
2. Observation: the absolute bias increases with the size of $T$.

**In the paper we prove slightly more general results for estimators that are not measurable.**

## Consistency

Intuitively, consistency can be guaranteed when $T$ is 'small' enough and if $f_t$ is 'smooth' enough. In the paper we have conditions for the general setting, but here we will – for brevity and simplicity – only discuss the case of importance sampling. In the following we assume that $\mathrm{E}^P(F) < +\infty$ and for some theorems $p_t$ are required to have an extension to values of $t$ outside of $T$. For the exact theorems, we refer to the paper.

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

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

## Example of No Consistency

We will look at an importance sampling estimator with central density $p = \frac{1}{2}\mathbb{I}_{[0,1]}$ and a countable set of densities

$$\left\{ \sum_{\ell=1}^{2k} a_\ell \mathbb{I}_{\left[\frac{\ell-1}{k}, \frac{\ell}{k}\right]} : k \in \mathbb{N} \text{ and } (a_1, \dots, a_{2k}) \text{ is a binary sequence with the same number of zeros and ones} \right\}.$$

For a sample of size $n$ it is always possible to choose a binary sequence of at most size $2n$, such that the corresponding density is zero on the sampled values.