Densities of almost-surely terminating probabilistic programs are differentiable almost everywhere

Carol Mak, C.-H. Luke Ong, Hugo Paquet, and Dominik Wagner
Department of Computer Science, University of Oxford, UK

Probabilistic Programs in Bayesian Inference

<table>
<thead>
<tr>
<th>Prior</th>
<th>Probalistic Language</th>
<th>Inference algorithm</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Why density?
Reason about correctness of inference algorithms

Why differentiation?
"Gradient-based" inference algorithms (HMC, stochastic VI) rely on derivatives exists "often enough"

Densities of almost-surely terminating probabilistic programs denote a distribution with a density that is not differentiable at some non-measure-zero set?

Q2: Can a probabilistic program denote a distribution with a density that is not differentiable at some non-measure-zero set?

Density of probabilistic program M gives the weight of a execution of M with trace s of all sampled values.

When is the density differentiable?

Failure of differentiability

Conditional
if(sample ≤ sample, score(1), score(0)), 1, [] →

Choice of primitives
if(f(sample) ≤ 0, score(1), score(0))
f is the differentiable function which is zero on fat Cantor set but strictly positive elsewhere

Termination
let _ ≜ N sample in score(1) [ [ s_i | s_i ∈ Q ] ]
N is a (det.) program that halts if input is rational

Stochastic symbolic execution (⇒) captures branching elegantly

The class \( \mathcal{F} \) of primitive functions is a set of partial, measurable functions \( \mathbb{R}^* \rightarrow \mathbb{R} \) including all constants and projections, 1) closed under composition and pairing; for all \( f \in \mathcal{F} \), 2) \( f \) is differentiable in the interior of its domain and 3) \( \text{Leb}(\partial f^\top([0,\infty))) = 0 \).

A probabilistic program M almost-surely terminate if the set of traces where M does not terminate has measure zero.

Densities of almost-surely terminating probabilistic programs are differentiable almost everywhere.

1. Identify the closure property for the class of primitive functions
2. Introduce stochastic symbolic execution
3. Deduce that deterministic programs denote almost everywhere differentiable functions

Hongseok Yang’s FSCD 2019 invited lecture

Analytic functions with co-domain \( \mathbb{R} \)

Analytic functions

when is the density differentiable?