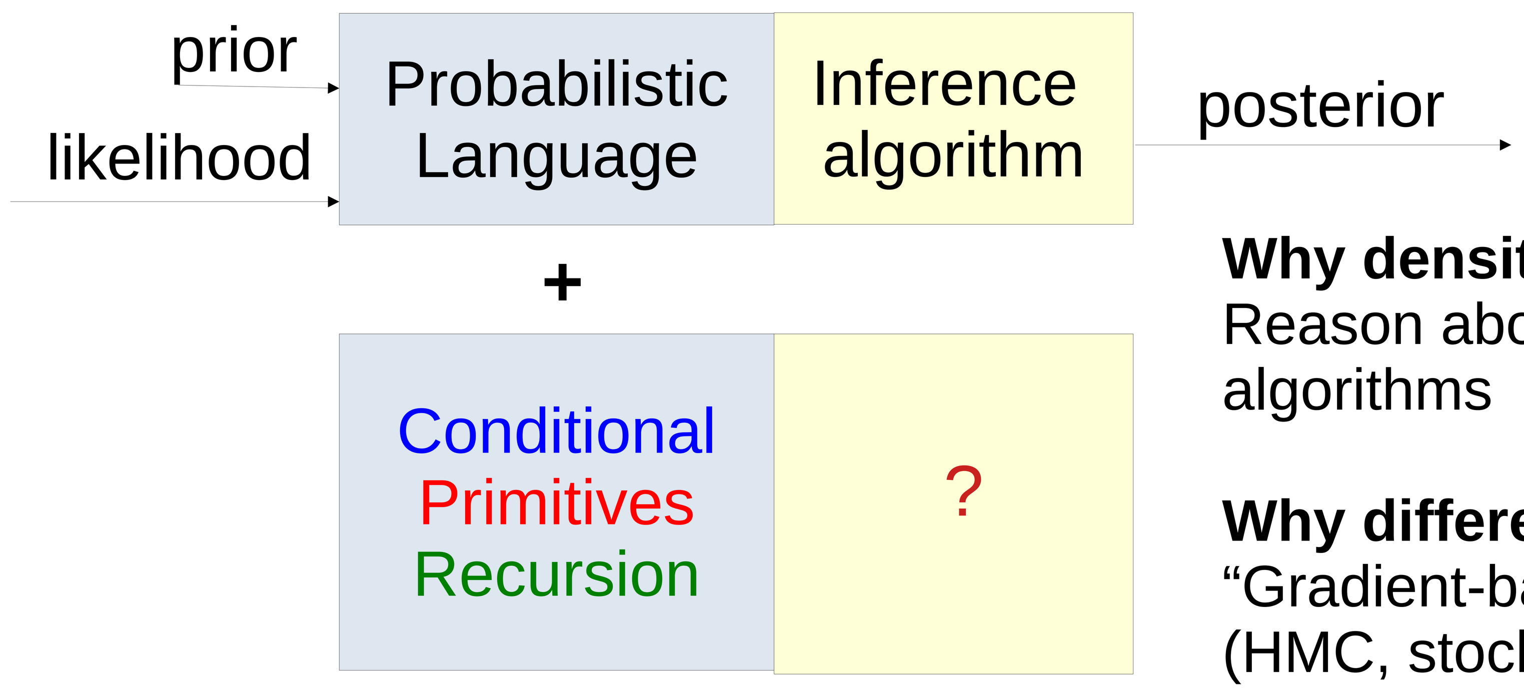
Densities of almost-surely terminating probabilistic programs are differentiable almost everywhere

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

Probabilistic Programs in Bayesian Inference



Why density?
Reason about correctness of inference algorithms

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

Hongseok Yang's FSCD 2019 invited lecture
Q2: Can a probabilistic program denote a distribution with a density that is not differentiable at some non-measure-zero set?

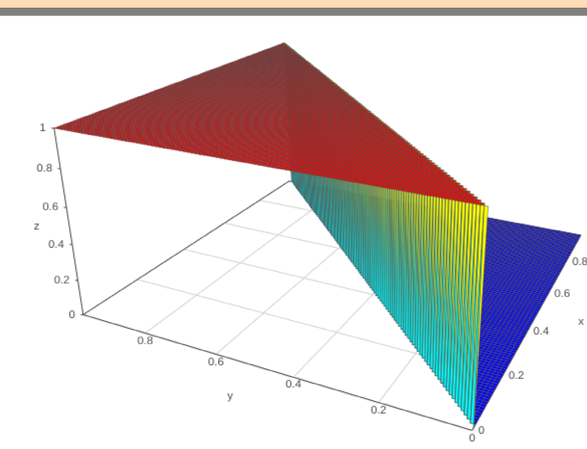
Density of probabilistic program *M* gives the weight of a execution of *M* with trace ***s*** of all sampled values.

When is the density differentiable?

Failure of differentiability

Conditional

if(sample ≤ sample, score(*1*), score(*0*))



Choice of primitives

if(*f*(sample) ≤ 0, score(*1*), score(*0*))

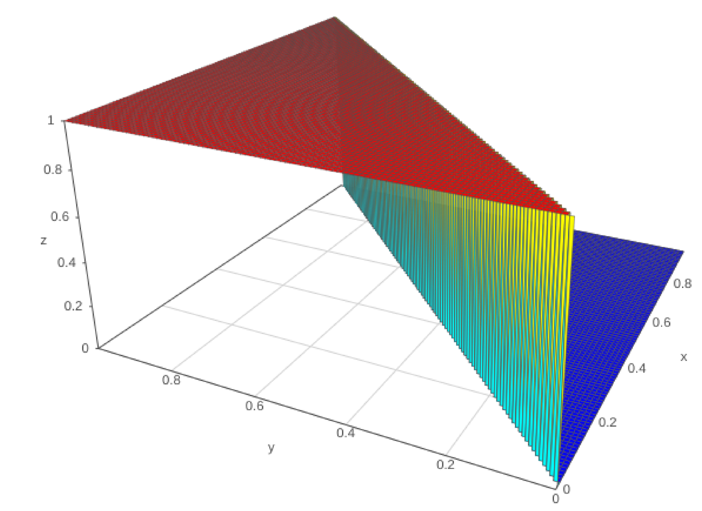
f is the differentiable function which is zero on fat Cantor set but strictly positive elsewhere

Termination

let _ = *N* sample in score(*1*) [{ [s_1] | $s_1 \in \mathbb{Q}$ }]

N is a (det.) program that halts if input is rational

⟨if(sample ≤ sample, score(*1*), score(*0*)), *1*, []⟩ $\to^*$
⟨if(0.6 ≤ 0.2, score(*1*), score(*0*)), *1*, [0.6,0.2]⟩ →
⟨score(*0*)), *1*, [0.6,0.2]⟩ →
⟨*0*, *0*, [0.6,0.2]⟩



Stochastic symbolic execution (⇒) captures branching elegantly

⟪if(sample ≤ sample, score(*1*), score(*0*)), λ[].1, { [] } ⟫
⇒ ⟪if($\alpha_1 \le \alpha_2$, score(*1*), score(*0*)), λ[s_1,s_2].1, $(0,1)^2$ ⟫

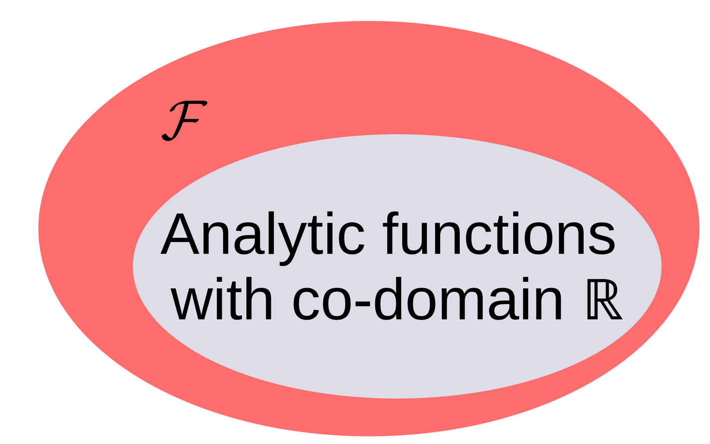
⇙ ⇘

⟪score(*1*), λ[s_1,s_2].1, { [s_1,s_2] | $s_1 \le s_2$ } ⟫
⇒ ⟪*1*, λ[s_1,s_2].1, { [s_1,s_2] | $s_1 \le s_2$ } ⟫

⟪score(*0*), λ[s_1,s_2].1, { [s_1,s_2] | $s_1 > s_2$ } ⟫
⇒ ⟪*0*, λ[s_1,s_2].0, { [s_1,s_2] | $s_1 > s_2$ } ⟫

Density is given ***naturally*** in stochastic symbolic execution!

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



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