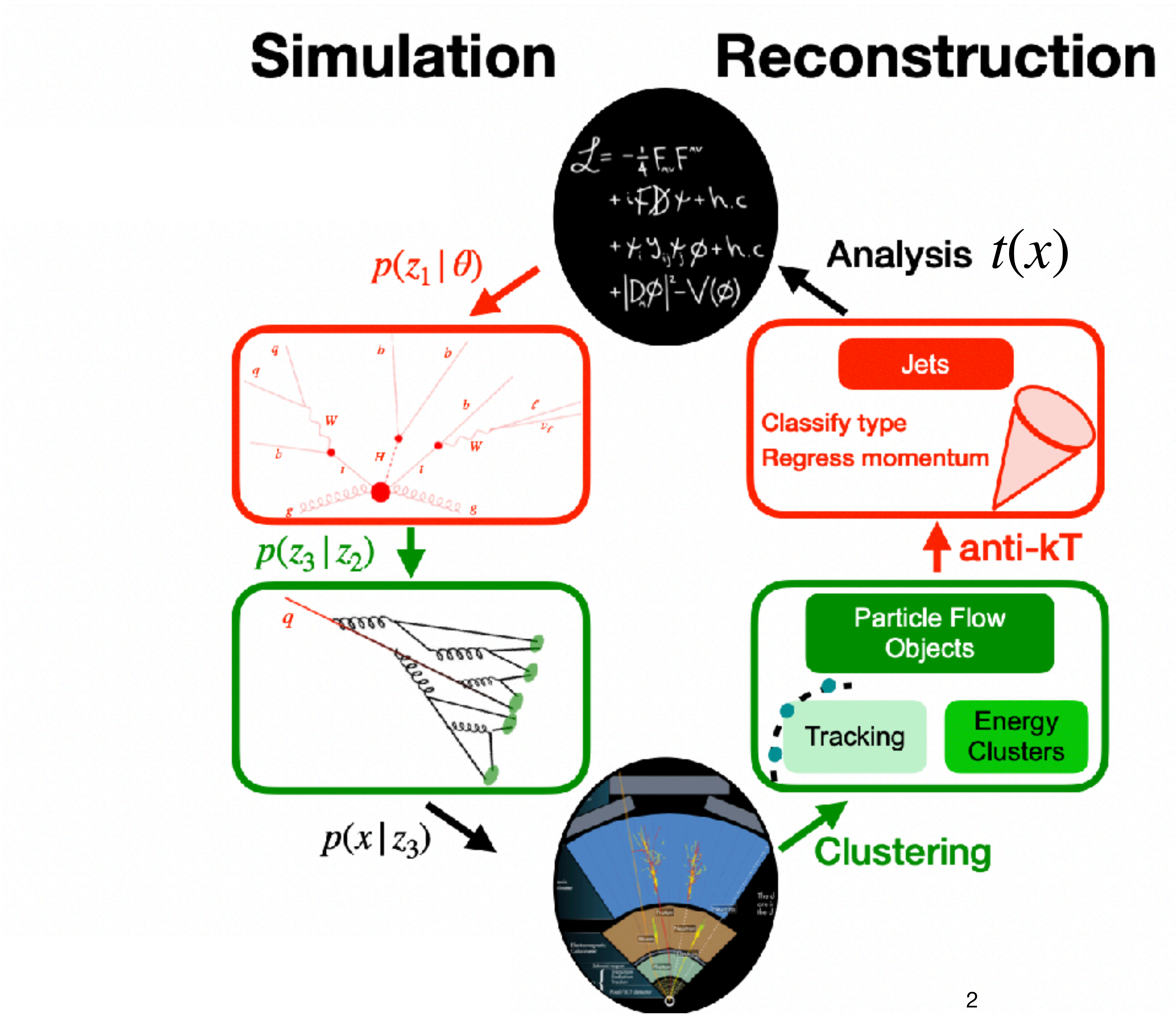
Differentiating discrete operations

Annalena Kofler, 05.03.2026

IRIS-HEP: Differentiable Analysis Blueprint

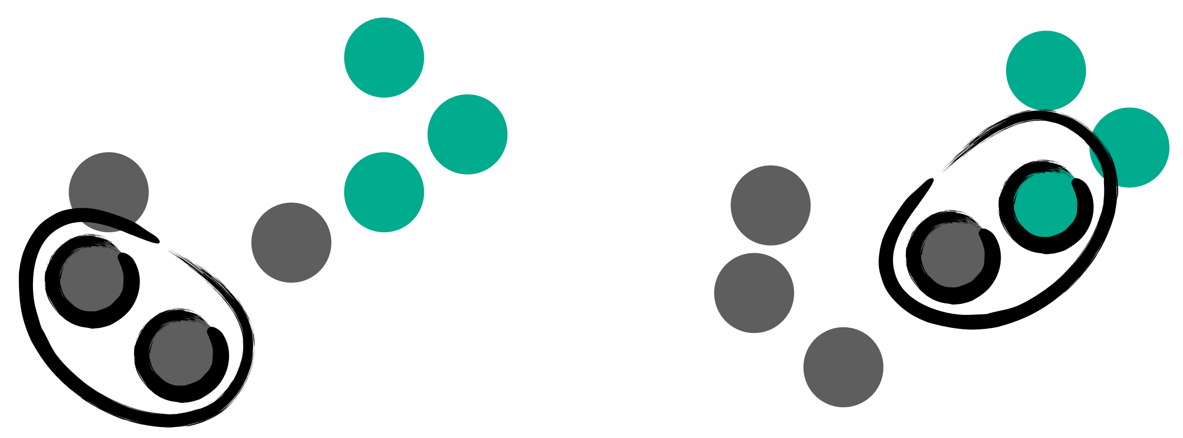
The end goal: A fully differentiable analysis pipeline



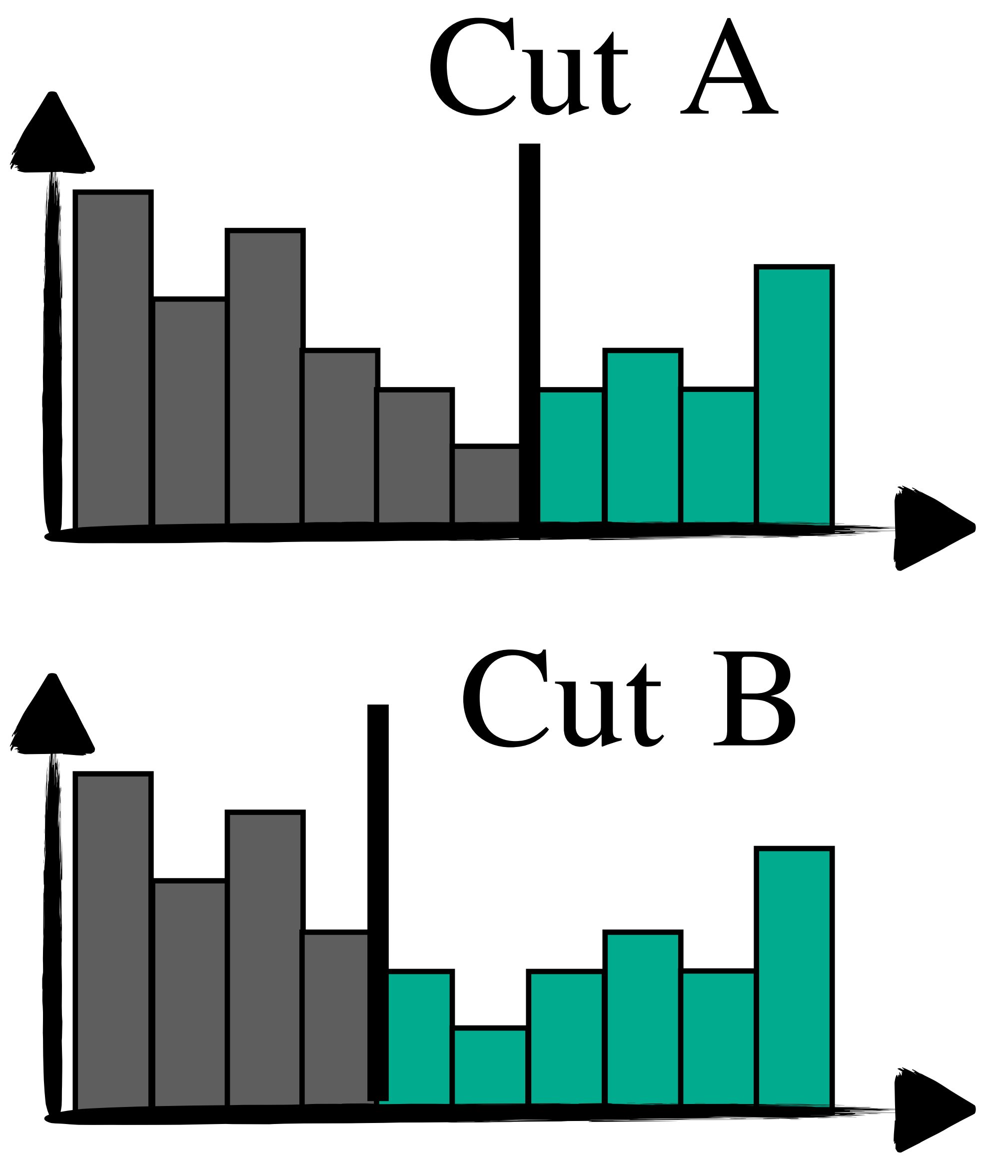
Slide adapted from K. Cranmer, H&N 2019
Slide adapted from N. Hartman, MIAPbP 2025
Slide adapted from G. Watts, MIAPbP 2025
Slide adapted from M. Algren, MIAPbP 2025

Pipeline includes **non-differentiable** operations per design!

Problem: discrete decisions



- Appears in many physics pipelines
- Examples:
  - Clustering: Which two points to cluster together?
  - Cuts: Which events to include which not?
  - …
- Common denominator: Gradient of discrete decisions



$$\nabla_\theta f(\theta) \quad \text{with} \quad f(\theta) = \begin{cases} \text{Cut A} & \text{if } \theta < 0 \\ \text{Cut B} & \text{if } \theta \geq 0 \end{cases}$$

In this talk:

How to make discrete decisions differentiable

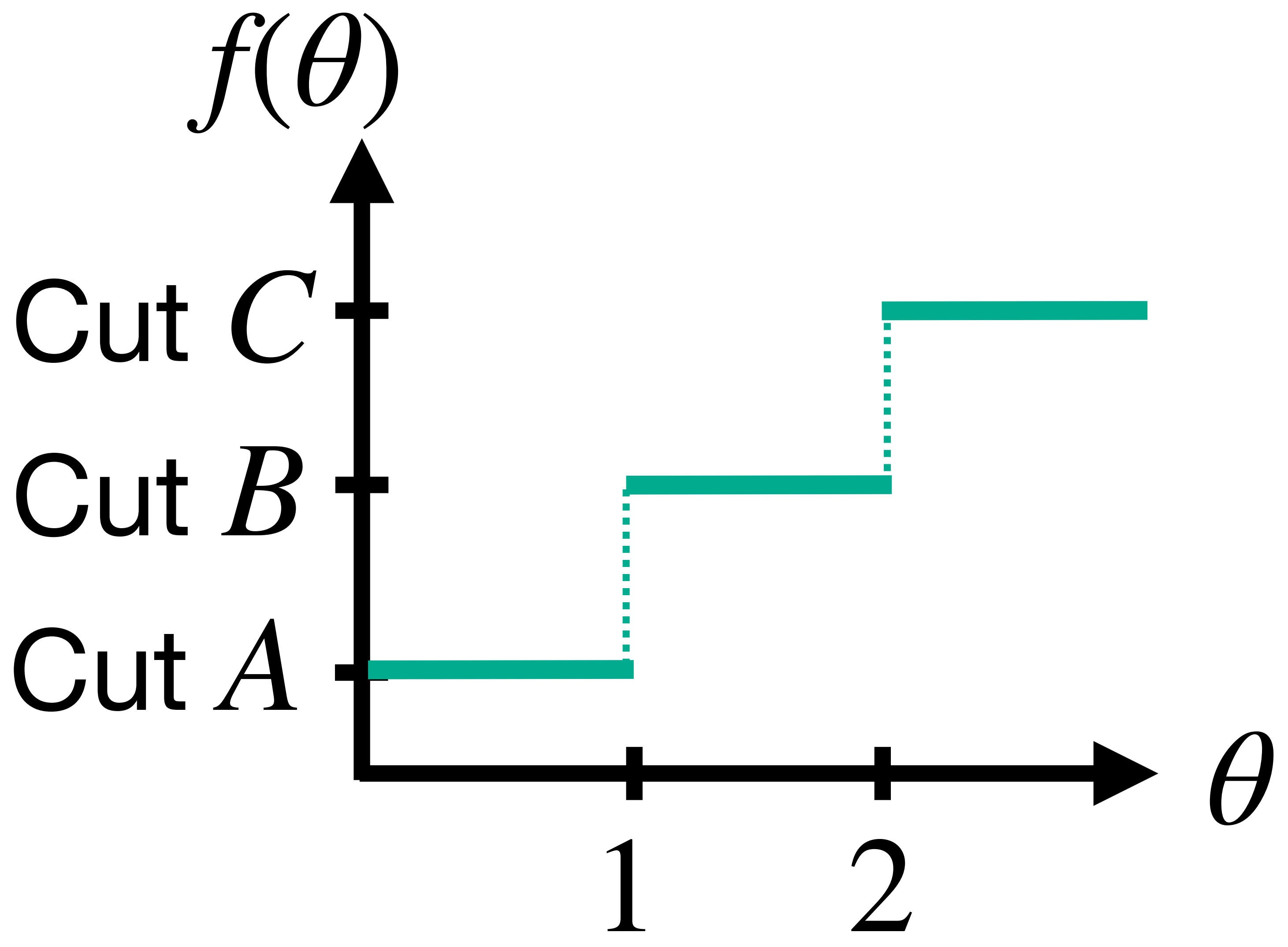
Applicable to many physics problems

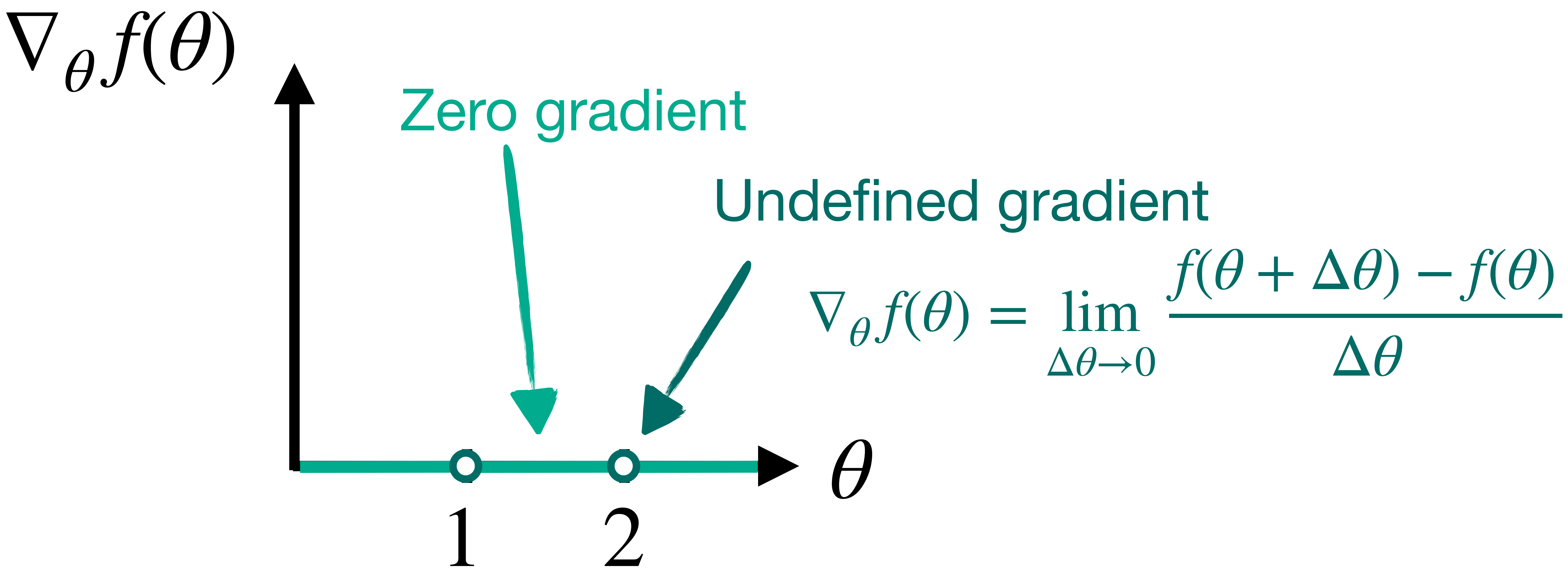
- Detector design in particle physics — Kagan & Heinrich, 2023, arXiv:2308.16680
- Heavy ion collision simulation — Magorsch & Heinrich, 2026, arXiv:2601.14399
- Reaction networks in biophysics — Burger, Kofler, et al., 2026, in preparation

Discrete decisions

Toy example with
3 options:

$$f(\theta) = \begin{cases} \text{Cut } A & \text{if } \theta < 1 \\ \text{Cut } B & \text{if } 1 \leq \theta \leq 2 \\ \text{Cut } C & \text{if } \theta > 2 \end{cases}$$





To get gradients for discrete decisions:

Step 1: Make the decision stochastic

Step 2: Make the stochastic decision differentiable

Step 1: Reinterpret decision as stochastic

Pick single best option deterministically → (Make Probabilistic) → Sample multiple good options stochastically

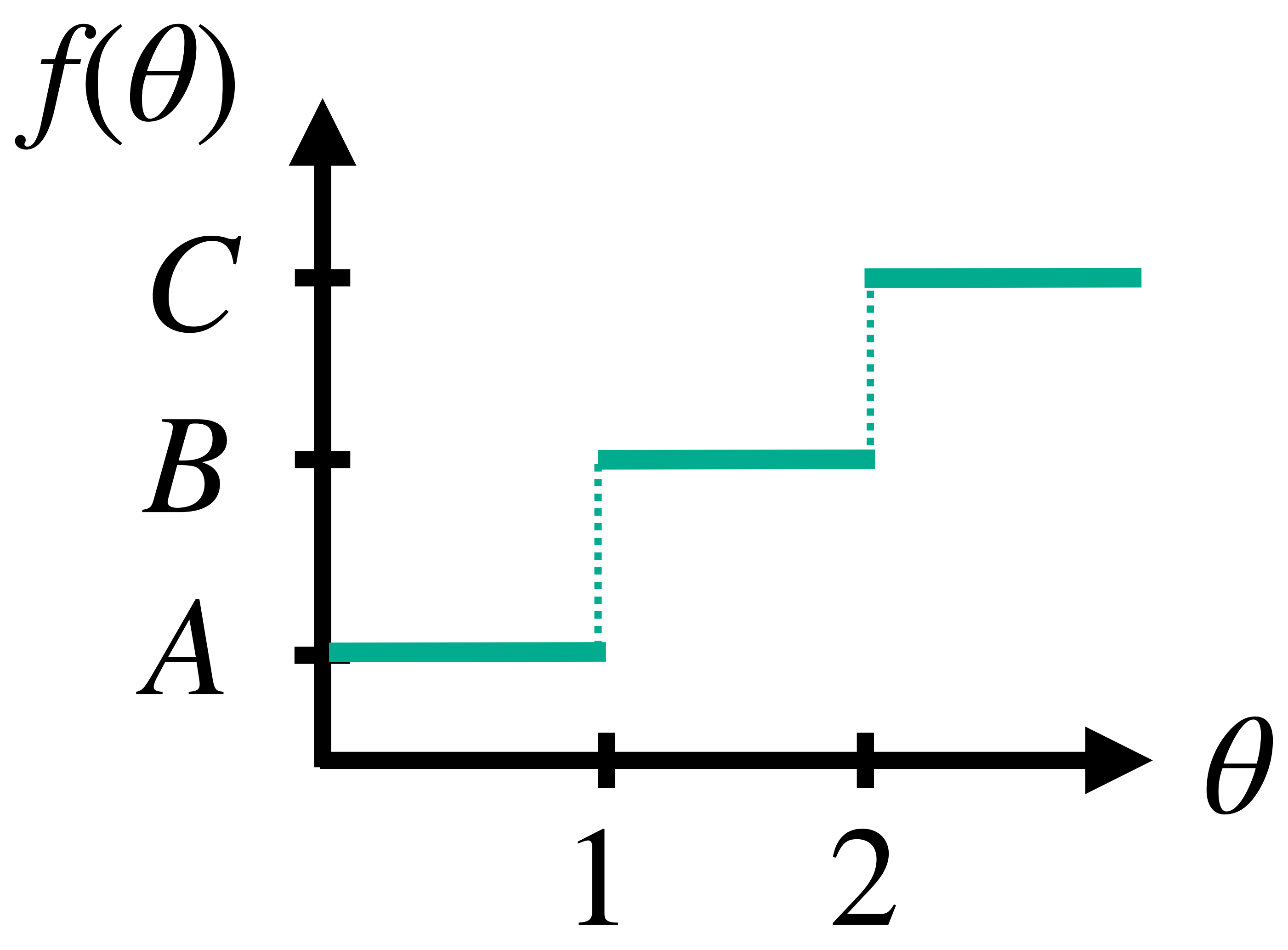
Deterministic	Stochastic
Single output	Multiple outputs → expectation value
$f(\theta)$	$f(\theta) = \mathbb{E}_{k \sim p_\theta(k)}[f(k)] = \sum_{k \in \{A,B,C\}} p_\theta(k) f(k)$

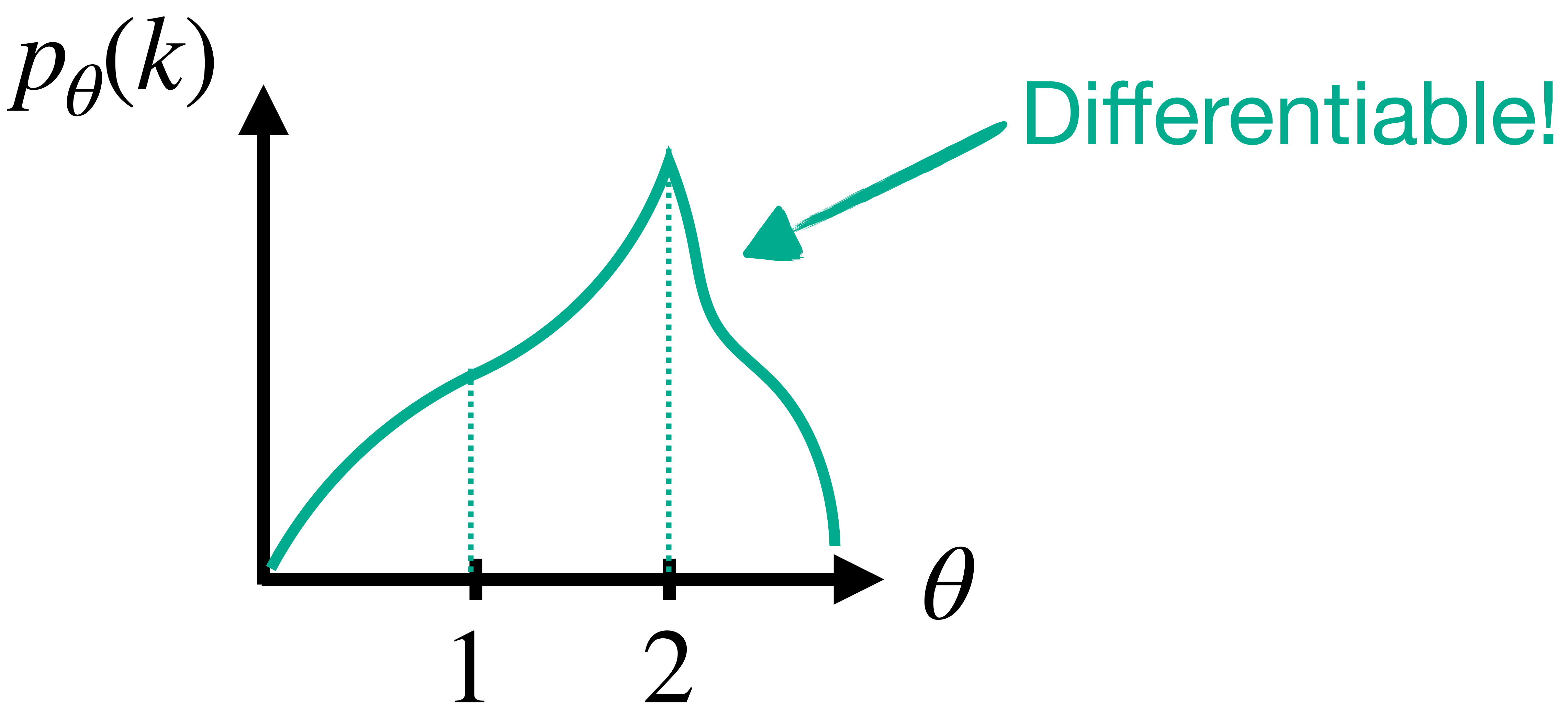
Why is a stochastic decision differentiable?

$$f(\theta) = \mathbb{E}_{k \sim p_\theta(k)}[f(k)] = \sum_{k \in \{A,B,C\}} p_\theta(k) f(k)$$

Define probability:

$$p_\theta(k) = \frac{\exp(l_k(\theta))}{\exp(l_A(\theta)) + \exp(l_B(\theta)) + \exp(l_C(\theta))}$$



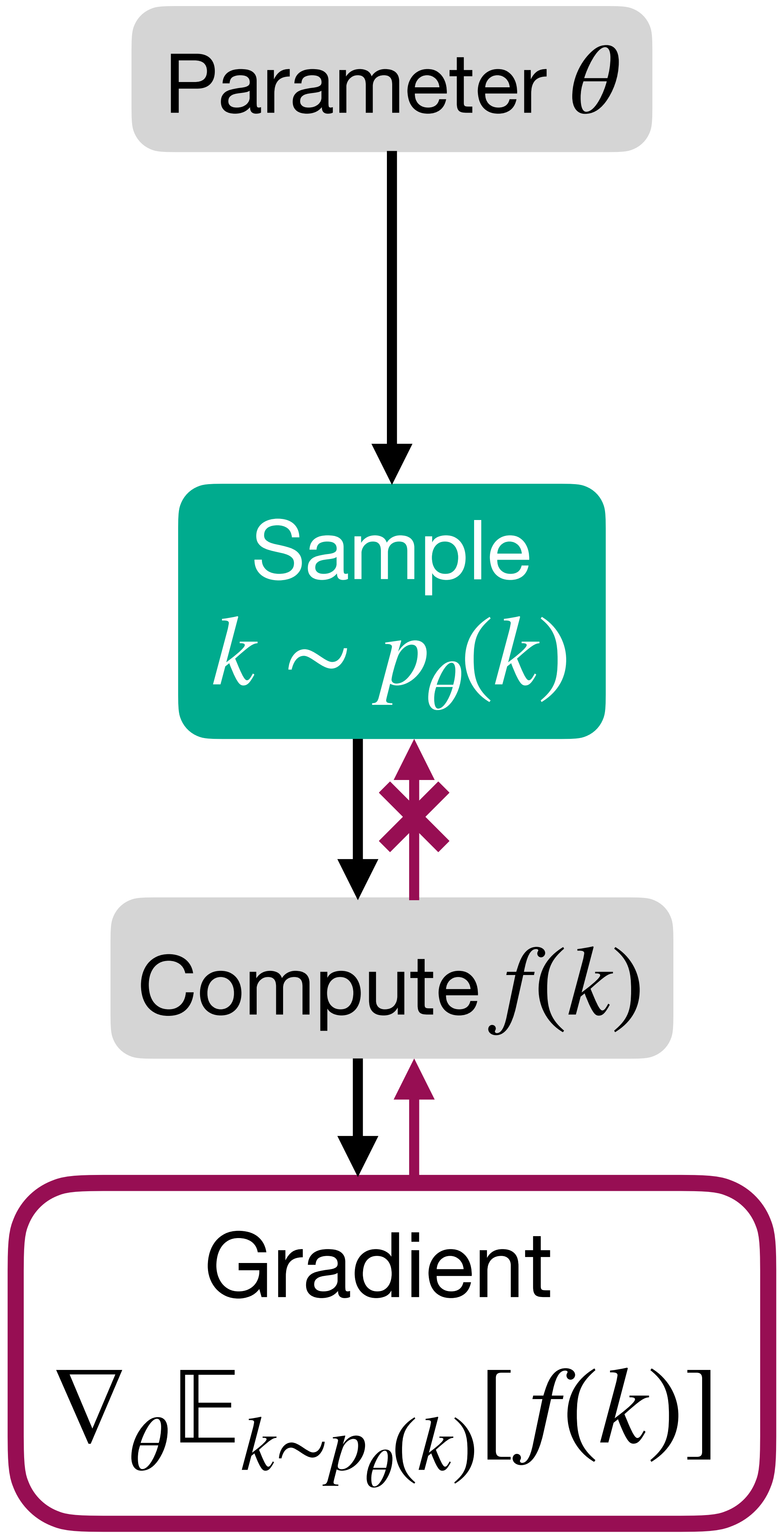


To get gradients for discrete decisions:

Step 1: Make the decision stochastic

Step 2: Make the stochastic decision differentiable

Problem



Score-based gradient estimator

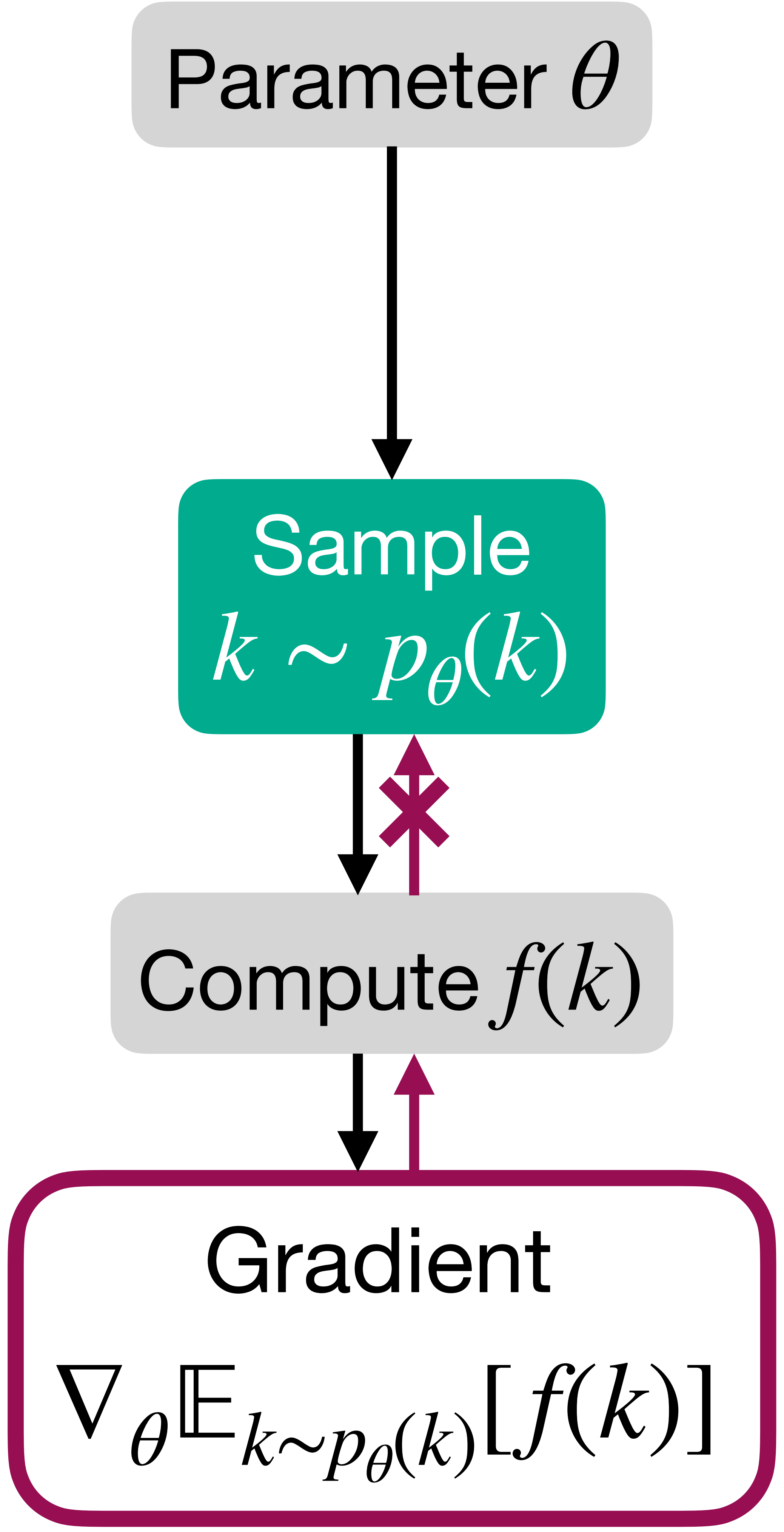
$$\nabla_\theta \mathbb{E}_{k \sim p_\theta(k)}[f(k)] = \sum_k f(k) \, \nabla_\theta p_\theta(k)$$

$$= \sum_k f(k) \, p_\theta(k) \, \nabla_\theta \log p_\theta(k)$$

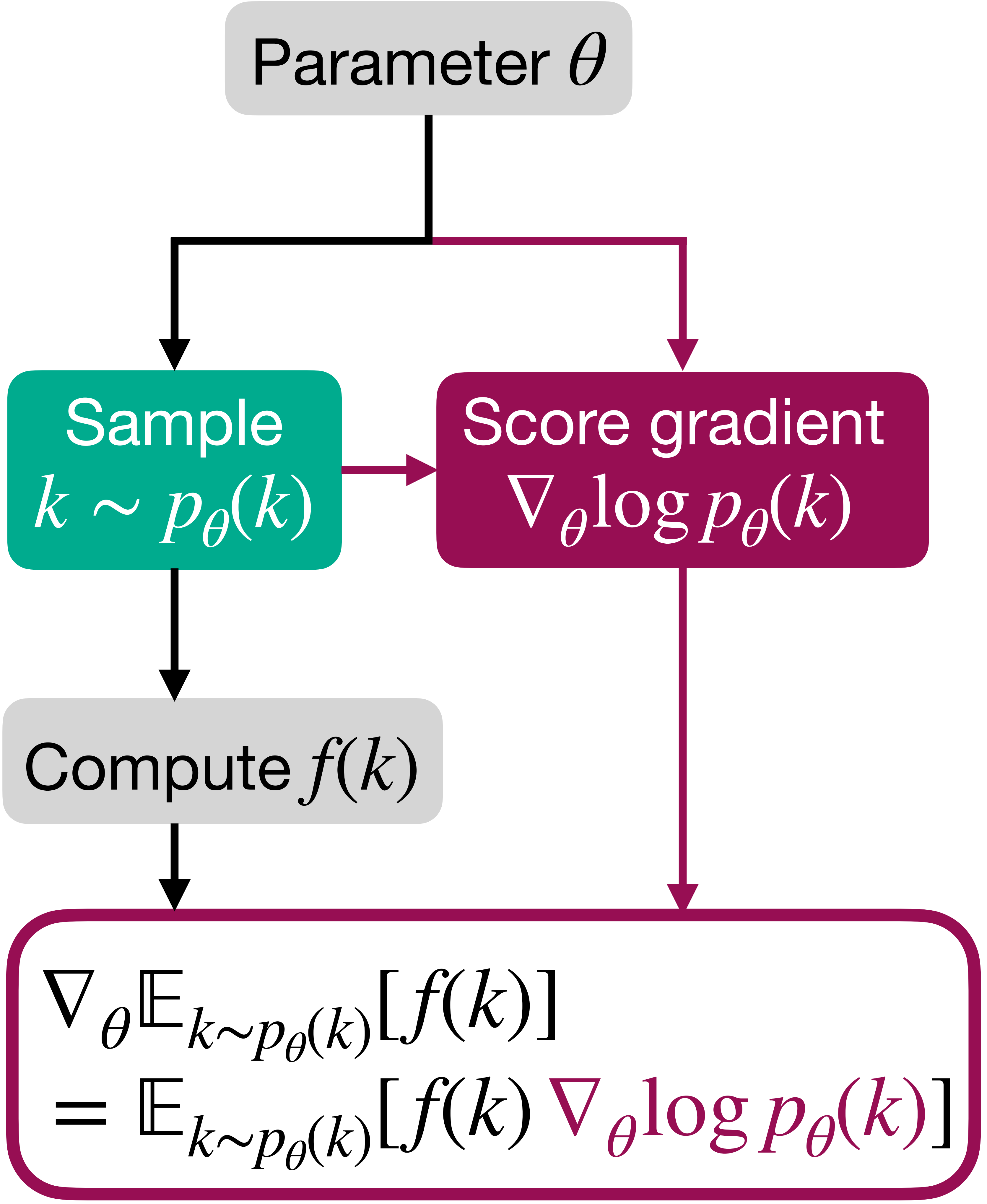
$$= \mathbb{E}_{k \sim p_\theta(k)}[f(k) \underbrace{\nabla_\theta \log p_\theta(k)}_{\text{Score}}]$$

Glynn 1990,
doi:10.1145/84537.84552

Problem

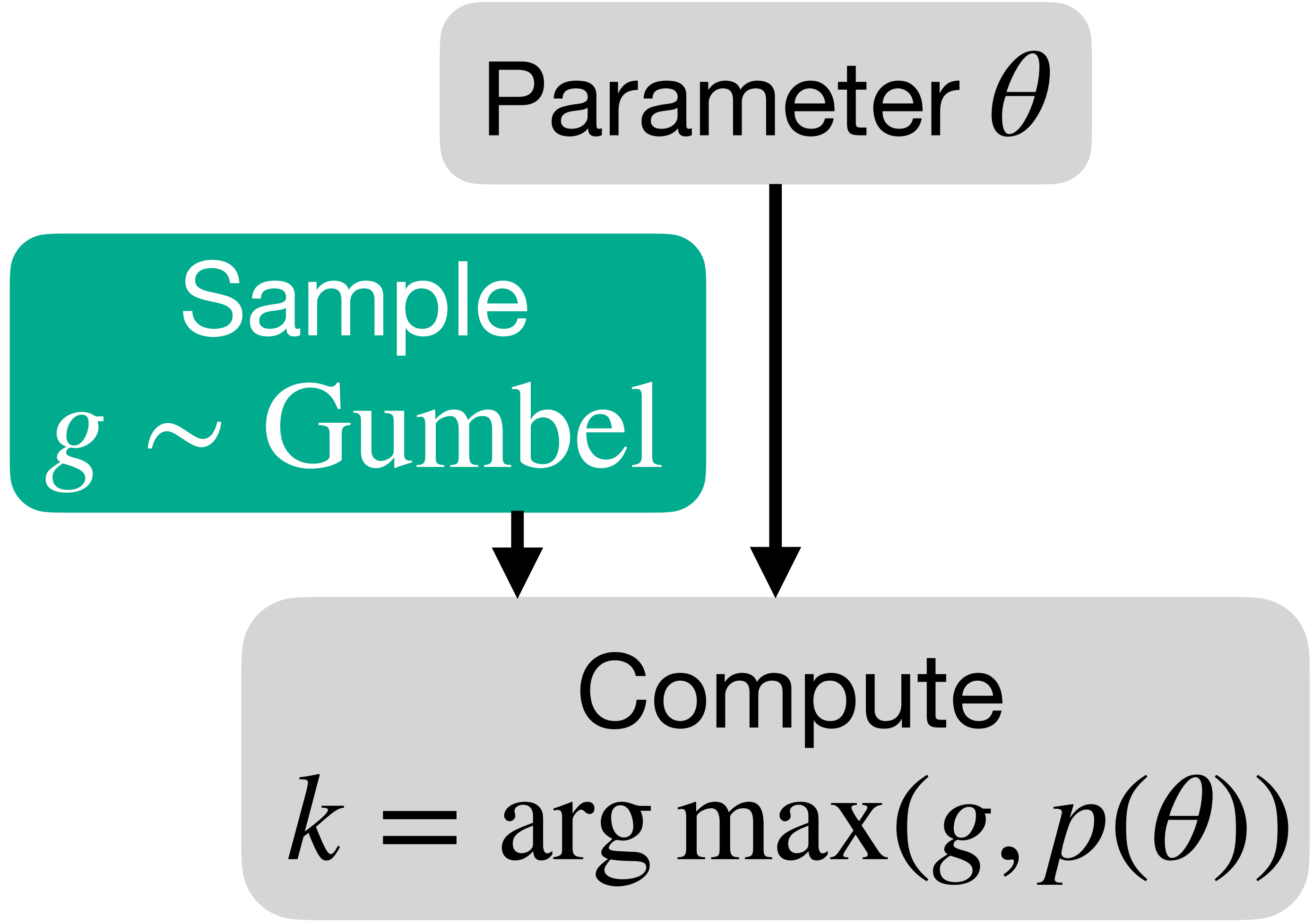


Score-based gradient estimator



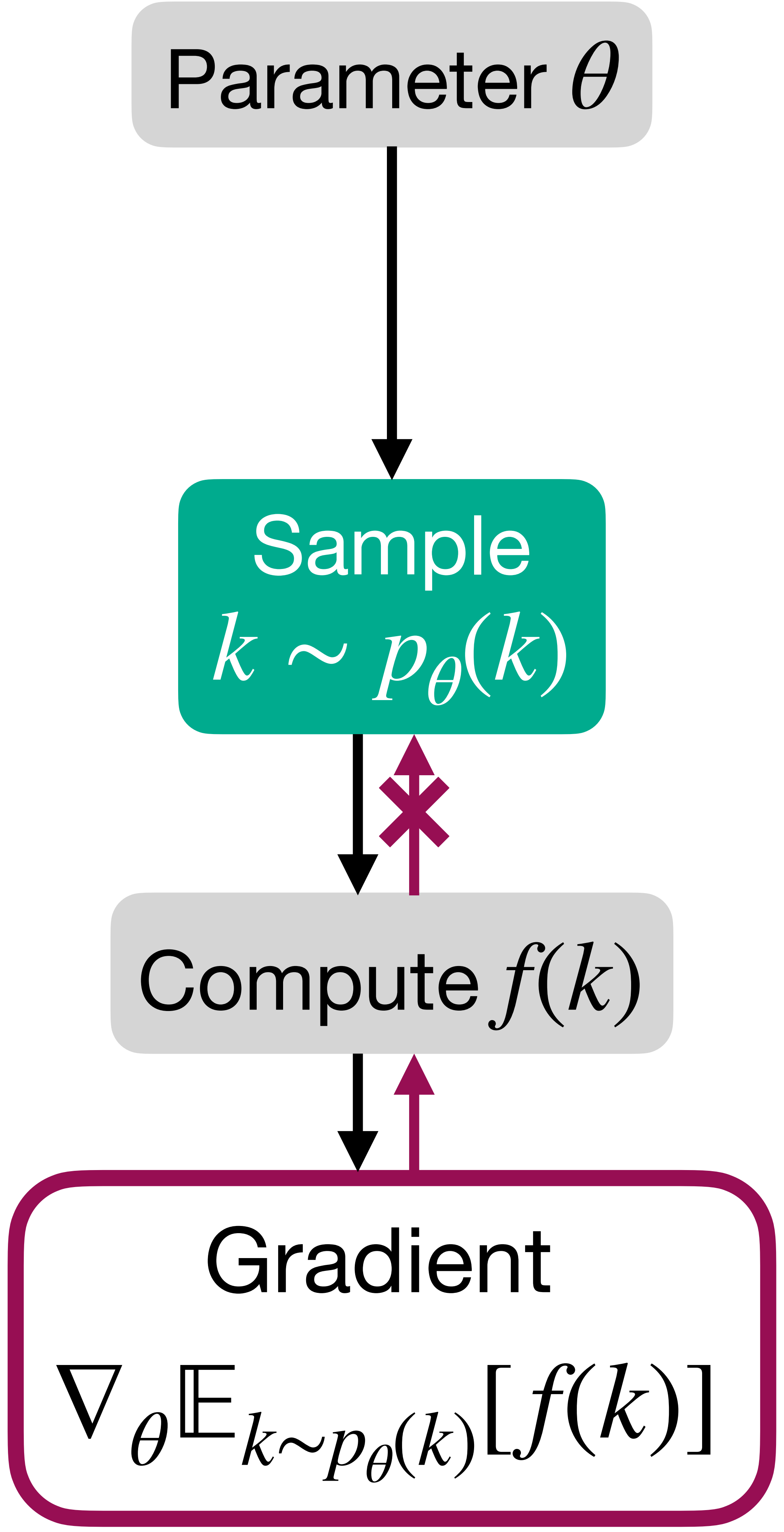
Glynn 1990,
doi:10.1145/84537.84552

Gumbel Softmax Straight-Through Estimator

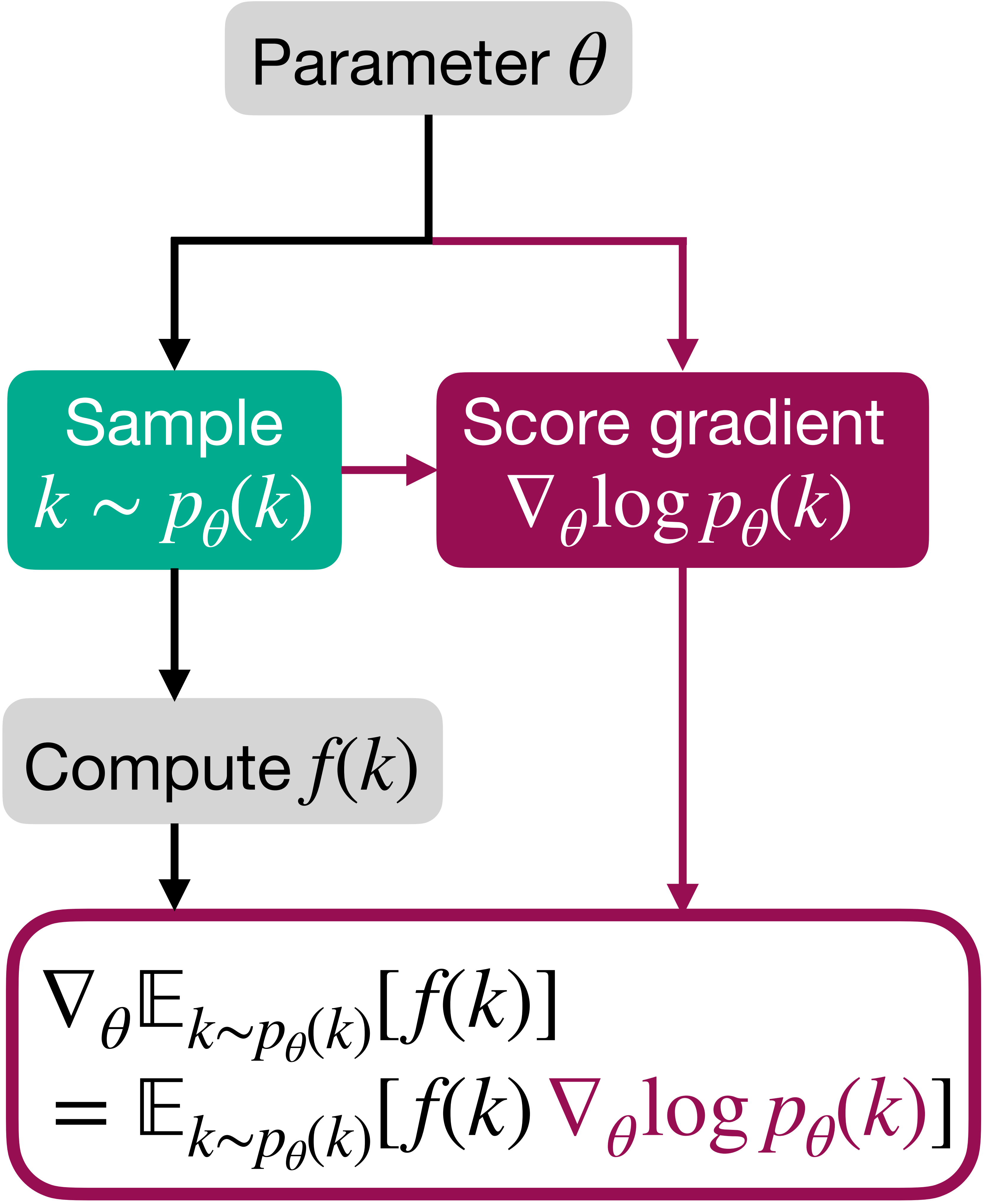


Jang et al. 2017,
arXiv:1611.01144

Problem

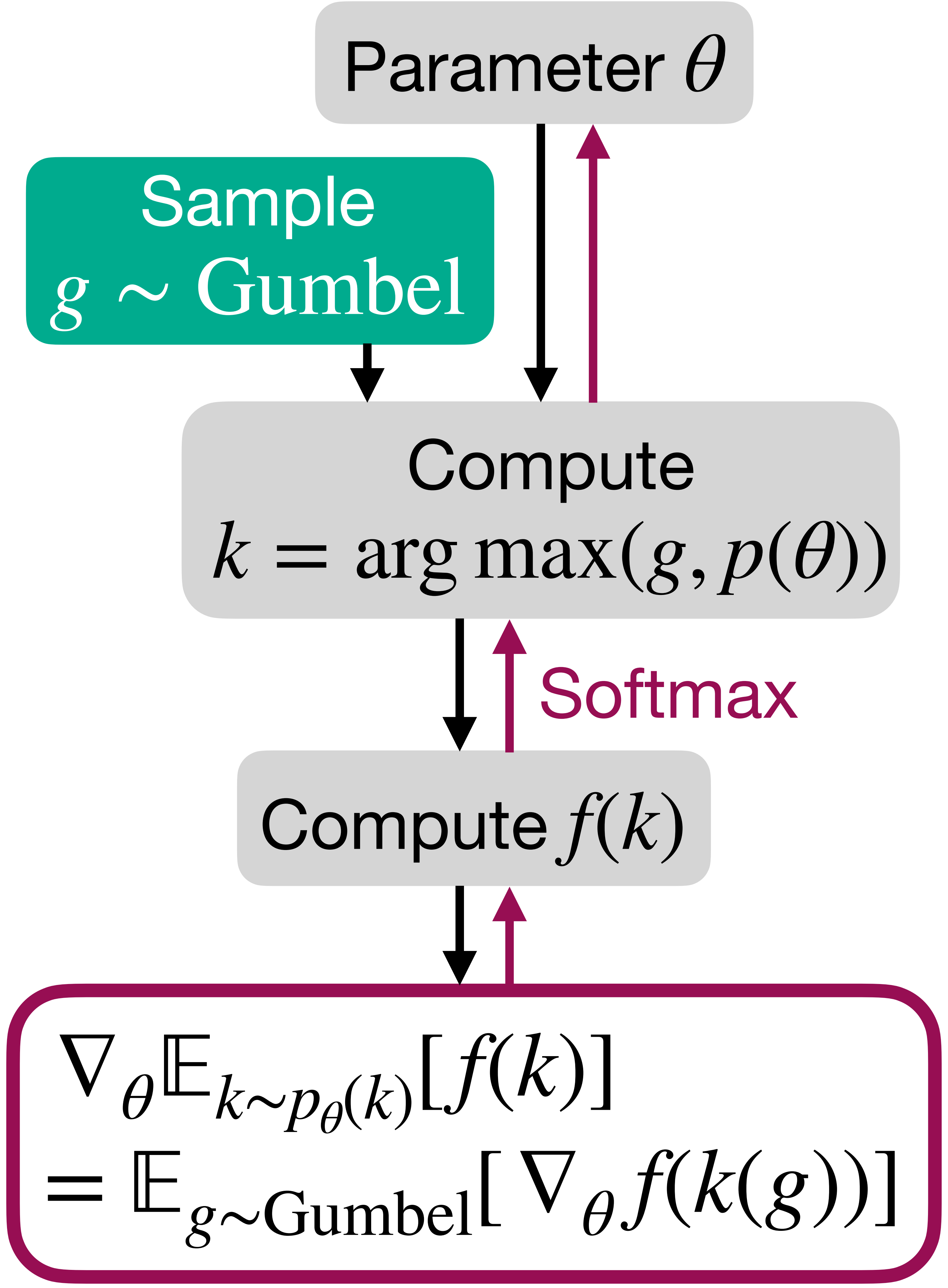


Score-based gradient estimator



Glynn 1990,
doi:10.1145/84537.84552

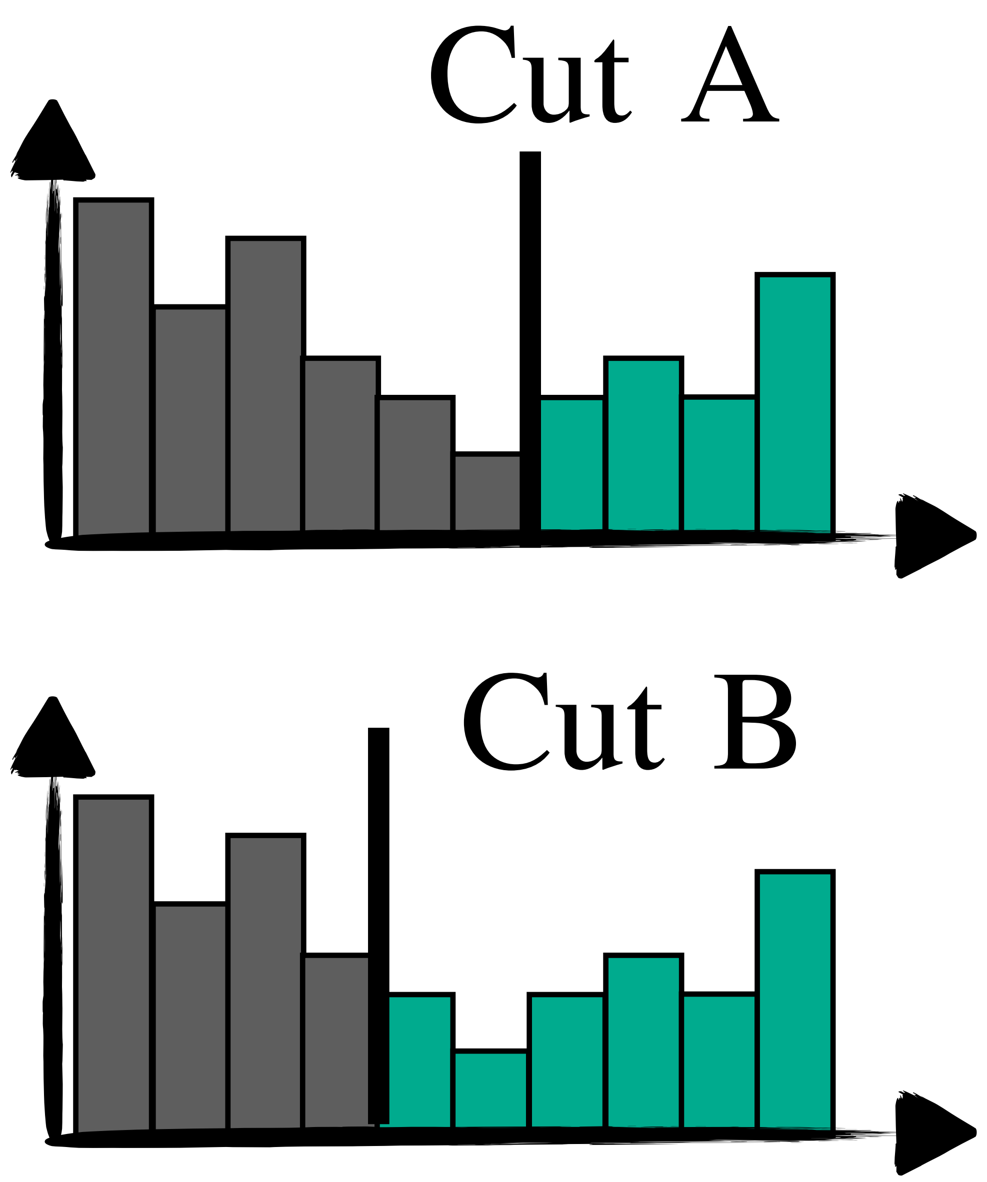
Gumbel Softmax Straight-Through Estimator



Jang et al. 2017,
arXiv:1611.01144

Take aways

- Discrete decisions not directly differentiable
- Solution:
  Step 1: Make it stochastic
  Step 2: Make it differentiable
- Gradient estimators:
  - Score based gradient
  - Gumbel-softmax straight-through estimator (GS-STE)
  - …



Questions?


THIS WORKSHOP:
GRADIENTS
GRADIENTS EVERYWHERE
imgflip.com