Probabilistic Programming Lecture #1: Introduction

Joost-Pieter Katoen

RWTH Lecture Series on Probabilistic Programming 2022-23

Overview

- 1 Introduction
- 2 Historical perspective
- 3 What is probabilistic programming?
- 4 Two striking examples
- **(5)** Why are probabilistic programs intricate?
- 6 What are we going to do in this course?
- What do we expect from you?

Joost-Pieter Katoen	Probabilistic Programming	1/76
Probabilistic Programming	Introduction	
Overview		
1 Introduction		
2 Historical perspective		
3 What is probabilistic programming		
Two striking examples		
5 Why are probabilistic programs in	tricate?	
6 What are we going to do in this c	ourse?	
 Introduction Historical perspective What is probabilistic programming Two striking examples Why are probabilistic programs into a striking to do in this c 	ç? tricate? ourse?	

What do we expect from you?

Joost-Pieter Katoen	Probabilistic Programming	2/76
Probabilistic Programming	Introduction	



Zoubin Gharahmani, Uber & University of Cambridge

REVIEW

Zoubin Ghahramani¹

Probabilistic machine learning

and artificial intelligence

Perspective

"There are several reasons why probabilistic programming could prove to be revolutionary for machine intelligence and scientific modelling."

[Ghahramani, Nature 2015]

Why? Probabilistic programming

 $1. \ldots$ obviates the need to manually provide inference methods

2. ... enables rapid prototyping

3. ... clearly separates the model and the inference procedures

Joost-Pieter Katoen	Probabilistic Programming	5/76
Probabilistic Programming	Historical perspective	
Overview		
1 Introduction		
2 Historical perspective		
3 What is probabilistic program		
Two striking examples		
6 Why are probabilistic progra	ms intricate?	
6 What are we going to do in	this course?	
What do we expect from yo		

Introductior

doi:10.1038/nature14541



The origin of probability theory



<section-header><section-header><section-header><section-header><section-header><text><text><text><text><text>

Christiaan Huygens addendum to "Mathematische Oeffeningen" 1660^a ^aKees Verduin, A Short History of Probability and Statistics

How "proofs" looked like

II. VOORSTEL.

Historical perspective

Als ick gelijcke kans hebbe tot a of b of c, het is my soo veel weerdt als of ick (a+b+c)/3 hadde.

Om dit wederom te vinden, soo zy als vooren gestelt x voor de waerde van mijn kans. Soo moet ick dan x hebbende weder tot de selfde kansse konnen geraecken door rechtmaetig spel. Laet dit het spel zijn, dat ick [491] tegen 2 andere speele, insettende ieder van ons drien x, ende laet ick met den eenen dese voorwaerde maecken, dat soo hy het spel wint hy my sal geven b, ende ick b aen hem, soo ick het kome te winnen. Met den anderen laet ick dese voorwaerde maecken, dat hy het spel winnende my sal geven c, of ick aen hem c als ick het win. Het blijckt dat dit spel rechtmaetig is. Ende ick sal daer door gelijcke kans hebben, om b te hebben, te weeten, als het den eersten wint, of c, als het den tweeden wint, of 3x - b - c als ick het win; want dan treck ick 3x die ingeset zijn, en geve daer van aen den eenen b, aen den anderen c. Indien nu 3x - b - c gelijck waer aen a, so soude ick gelijcke kans hebben tot a of b of c. So stel ick dan 3x - b - c = a en komt x = (a+b+c)/3, voor de waerde van mijn kans. Op gelijcke manier werdt gevonden, dat als ick gelijcke kans hebbe tot a of b of c of d, dit soo veel weerdt is als (a+b+c+d)/4, ende soo voorts.

Joost-Pieter Katoen Probabilistic Programmin Joost-Pieter Katoer Historical perspective robabilistic Programm The origin of programming The Analytical Engine Sketch of Lovelace's program turned a complex The Analytical Engine formula into simple calculations that could **Invented by Charles Babbage** be encoded on punched cards and fed into Charles Babbage's Analytical Engine, a By L. F. MENABREA mechanical computer that he designed but of Turin, Officer of the Military Engineers never built. She published it in 1843, a from the Bibliothèque Universelle de Genève, October, 1842, No. 82 century before the modern computer age. With notes upon the Memoir by the Translator ADA AUGUSTA, COUNTESS OF LOVELACE Ada Lovelace Translated by Ada Lovelace, 1836^a ^aAndrew L. Russell and Robin Hammerman, Ada'a Legacy: Cultures of Computing from the Victorian to the Digital Age, 2015.

Development of probabilistic models







Probabilistic Program

Historical perspective

"I want to put in something about Bernoulli's Number, in one of my Notes, as an example of how an explicit function may be worked out by the engine, without having been worked out by human head and hands first."



Joost-Pieter Katoen

loost-Pieter Katoer

Probabilistic Program

How "programs" looked like

					Diagram for the o	cimp	atation	i her	the E	igine	of the	Num	lars a	f Be	seally	. See Note G. (pag	722 et 10	4J				
L							Deta									Working Variables			1	Reptile 7	-	
Summer of Operation	Nation of Operation	-	Karnahlas reschiling inschiling	Indication of change in the value on any Tretable	Associated of Residue	FOes	1000m N	\$109 a. +	5'000 C	50000 [5"Os a	20ees	Pase []	5000 D		***0 ²	Pane	*0***	Part of	F. 1.10.2	F. Mado	12 ===0.4
	* * * * *	17, ×17, 17, -17, 17, +17, 17, +17, 17, +17, 17, +17, 17, +17, 17, +17, 17, +17,	17.0 17.0 19.0 17.0 17.0 17.0 17.0 17.0 17.0 17.0 17		$\begin{array}{l} = T \\ = T \\ = T + 1 \\ = 2 + 1 \\ = 2 + 1 \\ T + 1 \\ = 1 \\ T + 1 \\ T + 1 \\ = 1 \\ T + 1 \\ T + 1 \\ = 1 \\ T + 1$			+ (+) + + +	15 51	2.	10	0.1.1	1.1.1	0 0 1 0				-1-1-1-4				
7.	- +	74-753 753 #873	нг <u>и —</u> ну —		= 2 + 0 = 2	1	-	•	-	1 1		-	-	-	+-1		-		-	-		
	- * + -	14, +113 14, 14, 14 14, 14, 14, 14 14, 14, 14, 14, 14, 14, 14, 14, 14, 14,	ти ти ти ти		$\begin{array}{l} = \frac{2\gamma}{T} = \lambda_{1} \\ = y_{1} + \frac{2\gamma}{T} = y_{1} \lambda_{1} \\ = -\frac{1}{2} \cdot \frac{2\gamma}{T+1} + y_{1} - \frac{2\gamma}{T} \\ = -\frac{1}{2} \cdot \frac{2\gamma-1}{T+1} + y_{1} - \frac{2\gamma}{T} \\ = z - T (=T) \end{array}$	111-	1111	0111	1 1 1	1 1 1 1	2. 1 1 1	1.1.1 4	1111			$\frac{\frac{1}{2}a}{\frac{1}{2}a} = b_1$	n.‡-n.	{-}, ==== (===)		i		
	C++3 ++	$ \frac{1}{1} 1$		Second Contents	$\begin{array}{c} c F s = 1 \\ c T s = T \\ c s = T \\ c s = T \\ c s = 1 \\ c s$	1 1 1 1 1 1 1	1.1.1.1.1.1.1	1111111	11111	1111111	11-1 11-1 11-1 11-1		2			$\left\{\frac{12}{T}, \frac{12-1}{T}, 12$						20111
	1	10 × 10 × 10 × 10 × 10 × 10 × 10 × 10 ×	Na	AWWW.	$\begin{array}{c} = \frac{2}{3} & \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ = \frac{2}{3} & \frac{2}{3} & \frac{2}{3} & \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ = \frac{2}{3} & +\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ = \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} \\ = \frac{1}{3} & \frac{1}{3} &$	111-	1. 0, 150	1.1 1.1	+ + + +	111.0	1 1 6 1	1 1 1 1 0	1111			· ·	**	$\left\{ A_{k} + B_{k} A_{k} + B_{k} A_{k} \right\}$	II.	2.	111	ALC: NO.
4 1	1.		h	110000	= B ₁ = + + 1 = 4 + 1 = 2	1.	1=		1	-	-		-	-	1-	-	1-	1	1-	1	-	-

Historical perspective

Lovelace's program to compute Bernoulli numbers

Joost-Pieter Katoen	Probabilistic Programming 13/76
Probabilistic Programming	Historical perspective

Its first program



Probabilistic Programming

¹Erik Verhagen. Geheugentrommels. 2016

Joost-Pieter Katoen

The first Dutch computer



Automatische Relais Rekenmachine Amsterdam-1 (ARRA-1), 1952

Joost-Pieter Katoen	Probabilistic Programming	14/76
Probabilistic Programming	Historical perspective	

Historical perspective



Rethinking the Bayesian approach



"In particular, the graphical model formalism that ushered in an era of rapid progress in AI has proven inadequate in the face of [these] new challenges.

A promising new approach that aims to bridge this gap is probabilistic programming, which marries probability theory, statistics and programming languages"

Daniel Roy, Computability, Inference and Modeling in Probabilistic Programming. 2011²

² MIT/EECS George M. Sprowls Doctoral Dis	ssertation Award	
Pieter Katoen	Probabilistic Programming	17/76

Probabilistic Programming

Joost-Pie

What is probabilistic programming?

What is probabilistic programming good for?

"The goal of probabilistic programming is to enable probabilistic modeling and machine learning to be accessible to the working programmer"



Andy Gordon



Tom Henzinger

Aditya Nori Raja

Sriram Rajamani

Probabilistic Programmin

Overview

1 Introduction

- 2 Historical perspective
- 3 What is probabilistic programming?
- 4 Two striking examples
- 5 Why are probabilistic programs intricate?
- 6 What are we going to do in this course?
- What do we expect from you?

Joost-Pieter Katoen Probabilistic Programming 18/70

Probabilistic Programming

What is probabilistic programming?

What is probabilistic programming good for?

"Why probabilistic programs? To provide a clear and high-level, but complete, language

for specifying complex [probabilistic] models"



The PPL landscape

Python	PyMC, Edward, TCP
Javascript	webPPL
C++	STAN
Scala	Figaro
С	Probabilistic-C
Excel	Fabular
R	R2

Every programming language has a probabilistic variant

Joost-Pieter Katoen	Probabilistic Programming	21/76
Probabilistic Programming	What is probabilistic programming?	

Conditioning = learning



Probabilistic Programming

Probabilistic program:

random sampling, conditioning, and usual control-flow constructs

What is probabilistic programming

What is probabilistic programming

Analysis:

sophisticated Markov chain Monte Carlo simulation approaches³

It is a young field: first PROBPROG conference in 2018

³typically based on rejection sampling and random walks Joost-Pieter Katoen Probabilistic Programmin

22/7

Probabilistic Programming

Bayesian learning



Q: What is the probability that initially there was a piranha? A: $^{2}/_{3}$

Applications





Probabilistic Programmin

Probabilistic Programming

What is probabilistic programming?

Famous randomised algorithms

- Randomised Quicksort (196X)
- Rabin-Miller's Primality Test (1980)
- Freivald's Matrix Multiplication (1977)
- Lehmann-Rabin's Randomised Mutual Exclusion (1981)
- Herman's Randomised Self-Stabilising Algorithm (1990)
- ► The Coupon's Collector Algorithm
- Skip Lists (1990) and Bloom Filters (1970)

Randomised algorithms



What? Some decisions are based on coin flips

► Why?

- Their conceptual simplicity
- Their speed
 - mostly faster than their deterministic counterpart
 - no particular input elicits worst-case behaviour
- Their existence
 - many solve problems that have no deterministic solution
- Types

Joost-Pieter Katoer



- 1. Las Vegas: always produces correct results, random runtime
- 2. Monte Carlo: may produce wrong results, deterministic runtime



What is probabilistic programming?

Probabilistic Programmi

What is probabilistic programmi

26/76

Probabilistic Programming

Las Vegas: Sorting by flipping coins



27/7

Monte Carlo: Matrix multiplication

Input: three N^2 square matrices A, B, and C

Output: yes, if $A \cdot B = C$; no, otherwise

Time complexity over the years:

- until end 1960s: cubic (= 3)
- ► 1969: 2.808
- ▶ 1978: 2.**796**
- ► 1979: 2.780
- ▶ 1981: 2.**522**
- 1984: 2.496
- ▶ 1989: 2.**376**
- 2014: 2.373
- ▶ 2100:

Joost-Pieter Katoen	Probabilistic Programming	29/76
Probabilistic Programming	What is probabilistic programming?	

Security



Turing Award Winners 2013

"Goldwasser and Micali proved (1982) that encryption schemes must be **random** rather than deterministic [...] an insight that revolutionised the study of encryption and laid the foundation for the theory of cryptographic security."

used in almost all communication protocols, Internet transactions and cloud computing

Monte Carlo: Freivald's matrix multiplication

Input: three $\mathcal{O}(N^2)$ square matrices A, B, and C Output: yes, if $A \times B = C$; no, otherwise



Deterministic: compute $A \times B$ and compare with C Complexity: in $\mathcal{O}(N^3)$, best known complexity $\mathcal{O}(N^{2.37})$

Randomised: 1. take a random bit-vector \vec{x} of size N 2. compute $A \times (B\vec{x}) - C\vec{x}$ 3. output yes if this yields the null vector; no otherwise 4. repeat these steps k times

Complexity: in $\mathcal{O}(k \cdot N^2)$, with false positive with probability $\leq 2^{-k}$

Joost-Pieter Katoen

Probabilistic Programming

What is probabilistic programming

What is probabilistic programmi

30/76

Probabilistic Programming

A Birthday Attack

Let crypto-hash function $h: \{0, 1\}^m \rightarrow \{0, 1\}^n$ with $Pr[h(x) = y] = 2^{-n}$

```
b := true;
i := 1;
while (b) {
  for j = 1 to m do // sample m bits uniformly
    xi[j] := uniform(1) // the j-th bit of xi
  for k = 1 to i-1 do
    b := b and h(xi) != h(xk)
    od;
    i++
}
```

How many iterations are needed to terminate with probability > 1/2?

[Katz & Lindall, Intro to Modern Cryptography, 2014]

The famous RSA-OAEP protocol

Oracle $\operatorname{Enc}_{pk}(m)$: $r \notin \{0,1\}^{k_0};$ $s \leftarrow G(r) \oplus (m \parallel 0^{k_1});$ $t \leftarrow H(s) \oplus r;$ return $f_{pk}(s \parallel t)$	Game IND-CCA2 : $(sk, pk) \leftarrow \mathcal{KG}();$ $(m_0, m_1, \sigma) \leftarrow \mathcal{A}_1(pk);$ $b \notin \{0, 1\};$ $c^* \leftarrow \operatorname{Enc}(pk, m_b);$
Oracle $\text{Dec}_{sk}(c)$: $(s,t) \leftarrow f_{sk}^{-1}(c);$ $r \leftarrow t \oplus H(s);$ if $[s \oplus G(r)]_{k_1} = 0^{k_1}$ then return $[s \oplus G(r)]^n$ else return \perp	$b' \leftarrow \mathcal{A}_2(pk, c^*, \sigma);$ return $b = b'$ Game POW : $(sk, pk) \leftarrow \mathcal{KG}();$
Oracle $G(x)$: if $x \notin \text{dom}(L_G)$ then $L_G[x] \notin \{0, 1\}^{n+k_1}$; return $L_G[x]$	$y \underset{z \in \{0, 1\}^{n+k_1}}{y \underset{z \in \{0, 1\}^{k_0};}{z \in \{0, 1\}^{k_0};}}$ $y' \leftarrow \mathcal{I}(f_{pk}(y z));$
Oracle $H(x)$: if $x \notin \text{dom}(L_H)$ then $L_H[x] \notin \{0,1\}^{k_0}$; return $L_H[x]$ 42	
-Pieter Katoen Probabi	ilistic Programming

Probabilistic Programming

What is probabilistic programming?

Probabilistic graphical models





Correctness

For every IND-CCA2 adversary A there exists an inverter I for which

$$|Pr_{\text{IND-CCA2}}[b = b'] - 1/2| \leq \underbrace{Pr_{\text{POW}}[y = y']}_{\text{depends on } I} + \dots$$



It took about 20 years to prove a tight upper bound Proof-checked by Barthe *et. al.* (2012)

 Joost-Pieter Katoen
 Probabilistic Programming
 34/76

 Probabilistic Programming
 What is probabilistic programming?

The importance of Bayesian networks

 $``\ensuremath{\mathsf{Bayesian}}\xspace$ networks are as important to AI and machine learning

as Boolean circuits are to computer science."

[Stuart Russell (Univ. of California, Berkeley), 2009]

Judea Pearl: The father of Bayesian networks





Turing Award 2011: "for fundamental contributions to AI through the development of a calculus for probabilistic and causal reasoning".

Joost-Pieter Katoen	Probabilistic Programming	37/76
Probabilistic Programming	What is probabilistic programming?	
Cognitive science		

Computational models that perform probabilistic inference over hierarchies of flexibly structured representations can address some of the deepest questions about the nature and origins of human thought.

[Tenenbaum, Kemp, Griffiths, Goodman, Science, 2011]

Probabilistic programming provides such modeling and inference

Printer troubleshooting in Windows 95



How likely is it that your print is garbled **given that** the ps-file is not and the page orientation is portrait?

[Ramanna et al., Emerging Paradigms in Machine Learning, 2013]

Joost-Pieter Katoen	Probabilistic Programming	38/76
Probabilistic Programming	What is probabilistic programming?	

Planning in Al: Robot navigation



Uncertainty: noisy sensors and actuators, unknown environment⁴

Probabilistic Programming

Joost-Pieter Katoen

39/76

Joost-Pieter Katoen

⁴Evans *et al.*, Modeling Agents with Probabilistic Programs, 2019

Probabilistic	Programming
TTODADIIIStic	riogramming

Overview

Introduction

2 Historical perspective

- 3 What is probabilistic programming?
- Two striking examples
- **5** Why are probabilistic programs intricate?
- 6 What are we going to do in this course?
- What do we expect from you?



The seismic localization problem



1996: Comprehensive Nuclear-Test-Ban Treaty (by 2019: 184 countries signed)

Joost-Pieter Katoen	Probabilistic Programming	41/76
Probabilistic Programming	Two striking examples	

International Monitoring System





Event location UN system



It is not that easy

Problems:

- ▶ UN system misses nearly one-third of all seismic events of interest
- ▶ About 50% of the reported events are spurious
- Large team of experts analyse automatic bulletins to improve accuracy

Sources of difficulty: Many sources of uncertainty

- > Travel time, attenuation of frequencies on Earth not accurately known
- Detectors are noisy (up to 90% of detections are false)
- Huge number of detections recorded per day

Use a probabilistic generative model. In fact: a probabilistic program.



Probabilistic Programming

Two striking examples



An infrasound station



Joost-Pieter Katoen Probabilistic Programming 46/76 Probabilistic Programming Two striking examples

Key question

Recall Bayes' rule:
$$Pr(X | Y = y) = \frac{Pr_{\theta}(X) \cdot Pr_{\phi}(Y = y | X)}{Pr(Y = y)}$$

To determine:

Joost-Pieter Katoen

$$x^* = \arg \max_{x} \underbrace{Pr_{\theta}(X = x)}_{\text{prior over events}} \cdot \underbrace{Pr_{\phi}(Y = y \mid X = x)}_{\text{how events propagate}}$$

and are measured

What is the most likely explanation for all sensor readings?

Probabilistic Programmin

The priors



Event rate distribution ($\lambda_e = 4.6$ per hour)





Event location log density



Detection probability

Joost-Pieter Katoen

Probabilistic Programming

Two striking examples

Probabilistic Programming

Event location revamped



A probabilistic program snippet in BLOG

#SeismicEvents ~ Poisson[T*λ]; Time(e) ~ Uniform(0,T) IsEarthQuake(e) ~ Bernoulli(.999); Location(e) ~ if IsEarthQuake(e) then SpatialPrior() else UniformEarthDistribution(); Depth(e) ~ if IsEarthQuake(e) then Uniform[0,700] else 0; Magnitude(e) ~ Exponential(log(10)); IsDetected(e,p,s) ~ Logistic[weights(s,p)](Magnitude(e), Depth(e), Distance(e,s)); #Detections(site = s) ~ Poisson[T*λ_i(s)]; #Detections(event=e, phase=p, station=s) = if IsDetected(e,p,s) then 1 else 0; OnsetTime(a,s) ~ if (event(a) = null) then Uniform[0,T] else Time(event(a)) + GeoTravelTime(Distance(event(a),s),Depth(event(a)),phase(a)) + Laplace(µ(s), σ(s)) Amplitude(a,s) ~ If (event(a) = null) then NoiseAmplitudeDistribution(s) else AmplitudeModel(Magnitude(event(a)), Distance(event(a),s),Depth(event(a)),phase(a)) Azimuth(a,s) ~ If (event(a) = null) then Uniform(0, 360) else GeoAzimuth(Location(event(a)),Depth(event(a)),phase(a),Site(s)) + Laplace($0,\sigma_a(s)$) Slowness(a,s) ~ If (event(a) = null) then Uniform(0,20) else GeoSlowness(Location(event(a)),Depth(event(a)),phase(a),Site(s)) + Laplace(0, \sigma_a(s)) ObservedPhase(a,s) ~ CategoricalPhaseModel(phase(a))

Arora, Russell, and Sudderth, Bull. of Seismological Society of America, 2013

eter Katoen	Probabilistic Programming	5

Probabilistic Programming

Joost-Pie

Two striking examples

A 🛄 nuclear test



February 12, 2013

Main findings

- ▶ 60% reduction of missed events compared to UN system No increase of the number of false alarms
- Found events missed by human analysis of UN system output
- ▶ 2014: UN announces to use the NETVISA⁵ system
- ▶ 2018: UN replaces their system by NETVISA

Since 2018, earthquakes and explosions (conventional and nuclear) are located and diagnosed by a probabilistic program⁶

⁶Mitchell Prize of American Statistical Assoc. & Int. Society for Bayesian Analysis Joost-Pieter Katoen Probabilistic Programmin

obabilistic Programming

Two striking examples

The SCENIC language

Aim: Use probabilistic programs to generate synthetic training data.⁷



Probabilistic Programming

Synthetic training data for deep learning



- Collecting real-world data is slow and costly: preprocessing+labelling
- ▶ Hard to observe and reproduce rare corner cases (e.g., a car accident)

Probabilistic Programm

Joost-Pieter Katoer

Two striking examples

Probabilistic Program

Two striking example

SCENIC example: A badly parked car



- 1 spot = OrientedPoint on visible curb
- 2 badAngle = Uniform(1.0, -1.0) * (10, 20) deg
- 3 Car left of (spot offset by -0.5 @ 0), \
- facing badAngle relative to roadDirection 4

⁵Network Processing Vertically Integrated Seismic Analysis

Syntax	Distribution
<pre>(low, high)</pre>	uniform on interval of \mathbb{R}
Uniform(value,)	uniform over given values
Discrete({value: wt,})	discrete with given weights
Normal(mean, stdDev)	normal with given μ , σ

Plus conditionals, loops, functions, methods and so on.

Joost-Pieter Katoen	Probabilistic Programming	57/76
Probabilistic Programming	Two striking examples	

SqueezeDet: Small, fast, and energy efficient



Convolutional NN for real-time object detection in autonomous driving⁸

Probabilistic Programming

⁸Wu, landola, Jin & Keutzer, CVPR 2017

Joost-Pieter Katoen

Example scenes



Three scenarios generated from a \pm 20-line $\rm SCENIC$ program



Baseline: Grand Theft Auto V







Hard SCENIC test case: Background error



Hard SCENIC test case: Missed object



Joost-Pieter Katoen	Probabilistic Programming 61/76
Probabilistic Programming	Two striking examples
	Two striking examples
– • • • •	

Precision and recall

Suppose a neural network for recognizing dogs in photographs identifies 8 dogs in a picture containing 12 dogs and some cats.

Of the 8 identified as dogs, 5 actually are dogs (true positives), while the rest are cats (false positives).

The program's precision is 5/8 while its recall is 5/12.

Precision is "how useful the vision recognition results are", and

Recall is "how complete the results are".

High precision: substantially more relevant results than irrelevant ones. High recall: algorithm returned most of the relevant results.

Joost-Pieter Katoen Probabilistic Programming 62/76 Probabilistic Programming Two striking examples Main findings

SCENIC enables generating more effective training sets

procision -	true pos	and	and	and recall -	true pos
precision -	true pos + false pos	anu	recall –	true pos + false neg	

Training set	Baseline test-set ⁹		"Hard" SCENIC test-set	
	precision recall		precision	recall
100/0 95/5	72.9 ± 3.7 73.1 ± 2.3	37.1 ± 2.1 37.0 ± 0.6	62.8 ± 6.1 68.9 ± 3.2	65.7 ± 4.0 67.3 ± 2.4

Test set results trained on 5,000 images from 100% screen shots or mixed with 5% "hard" SCENIC cases

Probabilistic Programming

 9 Screen dumps from playing the video game Grand Theft Auto V

63/76

Why are probabilistic programs intricate?

Overview

- Introduction
- 2 Historical perspective
- 3 What is probabilistic programming?
- 4 Two striking examples
- 5 Why are probabilistic programs intricate?
- 6 What are we going to do in this course?
- What do we expect from you?

Joost-Pieter Katoen	Probabilistic Programming	65/76
Probabilistic Programming	Why are probabilistic programs intricate?	

Issue 1: Program correctness

Classical programs:

- A program is correct with respect to a (formal) specification "for input array A, the output array B is sorted and contains all elements contained in A"
- Defines a deterministic input-output relation
- ▶ Partial correctness: if an output is produced, it is correct
- ▶ Total correctness: in addition, the program terminates

Probabilistic programs:

- They do not always generate the same output
- They generate a probability distribution over possible outputs

Why are probabilistic programs intricate

Three elementary issues

- 1. Program correctness
- 2. Termination
- 3. The runtime of a program

Joost-Pieter Katoen Probabilistic Programming 66/7

Issue 2: Termination

Probabilistic Programmi

Classical programs:

- They terminate (on a given/all inputs), or they do not
- ▶ If they terminate, they take finitely many steps to do so
- Showing program termination is undecidable (halting problem)

Probabilistic programs:

- ▶ They terminate (or not) with a certain likelihood
- They may have diverging runs whose likelihood is zero
- They may take infinitely many steps (on average) to terminate even if they terminate with probability one!
- Showing "probability-one" termination is "more" undecidable
 - > and showing they do in finite time on average, even more!

Why are probabilistic programs intricate

Issue 3: The program's runtime

Classical programs:

- They have a deterministic, fixed run-time for a given input
- Runtimes of terminating programs in sequence are compositional: if P and Q terminate in n and k steps, then P;Q halts in n+k steps
- Analysis techniques: recurrence equations, tree analysis, etc.

Probabilistic programs:

- Every runtime has a probability; their runtime is a distribution
- Runtimes of "probability-one" terminating programs may not sum up if P and Q terminate in n and k steps on average, then P;Q may need infinitely many steps on average
- Analysis techniques: involve reasoning about expected values etc.

post-Pieter Katoen	Probabilistic Programming	69/76
obabilistic Programming	What are we going to do in this course?	
This course's topics		
The webPPL Language	Conditioning	
Markov Chains	Recursion Theory	

- Markov Chain Monte Carlo
- pGCL: Syntax and Semantics
- Fixed Point Theory
- Weakest Preconditions
- Probabilistic WP
- Bounds and Invariants

- Recursion Theory
- Termination
- Proving AST
- Expected Runtimes
- Bayesian Networks
- Probabilistic Databases
- Automated Program Analysis

Overview

Introduction

- 5 Why are probabilistic programs intricate?
- 6 What are we going to do in this course?
- What do we expect from you?

What are we going to do in this course

Probabilistic Programm

Course material

Joost-Pieter Katoen

Probabilistic Programmin

- \blacktriangleright Lecture material = the slides + the lectures + recent papers
- Noah Goodman and Andreas Stuhlmüller: The Design and Implementation of Probabilistic Programming Languages, 2016. On-line available dippl.org
- Benjamin Kaminski:

Advanced Weakest Precondition Calculi for Probabilistic Programs, 2019.

Gilles Barthe, Joost-Pieter Katoen and Alexandra Silva (editors): Foundations of Probabilistic Programming, 2020.

Annabelle Mclver and Carroll Morgan: Abstraction, Refinement and Proof for Probabilistic Systems, 2005.

What are we going to do in this course?

Lectures

- ► Tue 16:30–18:00 (AH3)
- ▶ Thu 16:30–18:00 (AH2)
- ▶ Start: Oct 11, 2022
- Two lectures per week
- End: early January 2023
- Available material per lecture:
 - Slides prior to the lecture
 - Annotated slides after the lecture

Website

https://moves.rwth-aachen.de/teaching/ws-22-23 and RWTH Moodle

		70/70
Joost-Pieter Katoen	Probabilistic Programming	/3//6
Probabilistic Programming	What do we expect from you?	
Evercise classes		

- You are expected to hand in homework exercises
 - mostly on a weekly basis
 - working in groups of three students
 - ▶ a few practical exercises; mostly theoretical exercises
 - solutions discussed at the exercise classes

Start:

- First exercise series: Fri Oct 15
- First exercise class: Fri Oct 22 (AH1)
- Assistants:
 - Lutz Klinkenberg, Philipp Schroer, and Tobias Winkler

Overview

Introduction

- 2 Historical perspective
- 3 What is probabilistic programming?
- 4 Two striking examples
- 5 Why are probabilistic programs intricate?
- 6 What are we going to do in this course?
- What do we expect from you?

Joost-Pieter Katoen Probabilistic Programming 74/76 Probabilistic Programming What do we expect from you?

Examination

- Form: written exam
- **•** Exam qualification: \geq 50% of points in exercise series
- Registration for exam: Nov 15–Jan 15
- Exam dates:
 - Feb 16, 2023, 15:00–17:00
 - Mar 18, 2023, 12:30–14:30
- A hand-written "cheat sheet" of 1 double-sided A4 allowed
- How to raise your chances to pass the exam?
 - Exercise. Exercise. Exercise.

What do we expect from you?

Next lecture

Thursday Oct 13, 16:30

Joost-Pieter Katoen

Probabilistic Programming