

2

2.0

Computing difficult integrals using the law of large numbers

Def. 1

Goal: Compute
$$I = \int_{\mathbb{X}} \phi(x) \pi(x) dx$$

Monte Carlo Truncate a Law of Large Numbers (LLN) Methods: converging to I.

Example • Simulate independent $X_1, ..., X_n$ from π . (simple Monte Carlo): • Return $\widehat{I}_n = \frac{1}{n} \sum_{i=1}^n \phi(X_i)$.

Goals of this course

- Motivate why we need Monte Carlo (MC) methods.
- For the users of MC:
 - Using probabilistic programming languages.
 - Overview of the modern MC toolbox.
 We will iterate:
 - Problem we can't solve with what we know so far.
 - Methodology that solves this problem.
- How theory helps you write **correct** + efficient MC.
- For the developers of new MC methods:
 - implementation and analysis of state-of-the-art MC algorithms.
 - Interested in what you work on to guide choice of topics.

Logistics

- Course webpage: <u>http://tinyurl.com/starferret</u>
 - Check the syllabus for references, assessment, info on final project and more
- My office hour: after lecture?
- TA: Sohrab Salehi. Office hour TBA.

Logistics

- BRING LAPTOP IN CLASS! (Ask about windows users)
- Languages supported: Xtend/Java and Python
- Feel free to use other ones but strongly recommend learning a compiled language if serious about development of MC methods
- Install (use Piazza if encounter problems):
 - git (see https://software-carpentry.org/)
 - R
 - Oracle Java 8
 - PPLs:
 - Stan: http://mc-stan.org/
 - Blang: <u>https://www.stat.ubc.ca/~bouchard/blang/Blang_IDE.html</u>

Motivation: phylogenetic Ex. 2

- Topology: shape of the tree (discrete)
- Branch length: each edge has a parameter encoding how much evolution on the branch (continuous)



- Data: DNA at leaves (for example)
- Goal: reconstruct the tree (and more! ancestral sequences, models, ...)

Why phylo is important

• Signature left by evolution key to understand function



- Where do we come from?
- Beyond the tree of life:
 - cancer phylogenetics
 - trees of languages



Challenges

- Trees space:
 - high-dimensional,
 - combinatorial,
 - non-convex



- Uncertainty does not vanish in asymptotic regime of interest (sequence length does *not* go to infinity)
 - In many scientific applications, it is critical to quantify this uncertainty

Optimization and simulation

x : unknown y : data

Cases where this is advantageous...

$$x^{\star} \in \operatorname{argmax} p(y|x)$$



$$x^{(i)} \sim \frac{1}{Z} p(x) p(y|x)$$



I. Uncertainty over latent combinatorial structures

x = unknown **phylogenetic tree**

Question: is {human, monkey, lemur} a clade?



Uncertainty over latent combinatorial structures

x = phylogenetic tree



Clade posterior probability: 2/3



- These cases arise in various contexts:
 - stochastic processes/Bayesian non-parametrics,
 - high-dimensional models,
 - partially identifiable models, ...

In high dimension, optimization and sampling are profoundly different



Motivations, continued







Large scale random effect models



- e.g.: spatial and/or temporal data
- Models: Gaussian Markov random fields, cox processes, etc.

Ex. 3a

National Cancer Institute

Motivation, continued



- large networks
 - www, e.g. bitcoin transactions, link analysis, etc

Ex. 3b

- biological processes
- sparse exchangeable graph models (e.g. Caron and Fox, 2017)

Image credit: Ryan Rossi

Astrophysics: Estimating the Ex. 3c age and fate of the Universe

- Goals: finding the Universe's
 - age
 - density (=> faith)
- Data: Cosmic Microwave Background (CMB): remnants of Big Bang
 - Detailed map from the Planck satellite
- Age, Physical constants => known distribution on CMP
- Invert using Bayes' rule



More example:

Ex. 3d

- Computing volumes
- Physics
- Time series

Def. 4 Setup: large scale Bayesian or random effects models

• We are given a density known up to a normalization constant

$$\pi(x) = \frac{\gamma(x)}{Z}$$

- Example: $\pi(x) = \frac{\text{joint}(x, y)}{\text{evidence}(y)}$ |x : unknown y : data
- We want a law of large number $\frac{1}{N} \sum_{i=1}^{N} \varphi(X^{(i)}) \to \int \varphi(x) \pi(\mathrm{d}x) \text{ a.s.}$ 'test function'
- **Note:** We almost never care about the samples themselves!

Overview of the literature Note 5



Overview of the literature

'Naive' MCMC

- Deterministic start
- Apply kernel ad nauseam
- Burn-in, etc



Modern methods

- Sequential change of measure-based
- Replica-based methods

Def. 6

Basics: Simple Monte Carlo

Simple Monte Carlo:

- Simulate independent $X_1, ..., X_n$ from π .
- Return $\widehat{I}_n = \frac{1}{n} \sum_{i=1}^n \phi(X_i)$.

Simple example

- Consider d-dimensional iid random standard normal vectors
- What is the mean of the distance to the origin?
- Write as:

$$I = \int_{\mathbb{X}} f\left(x\right) dx$$

1.3 A Brief History of Monte Carlo M

Ex. 7

Computational motivation ^{Prop. 8}

Integrals: ID case $I = \int_{\mathbb{X}} f(x) dx$ 1.3 A Brief History of

1.3 A Brief History of Monte Carlo Methods Recall that, for out Monte Carlo \underline{m} ethod $\underline{h} \overset{n-1}{\underline{h}}$ confidence, interval was shr Recall that, for out Monte Carlo method the confidence interval with Brinkin Figure 1 atofate on the However, it is easy to see that its speed of convergence is of the same order, regardless of the dimension of the trapport of finistimenose the tease for interview interview of the terministic material of the terministic material of the team of the terministic material of the team of team ecall that differsions internet in Garlor method the confidence interval was shrinkings two-dimensional function f the error made by the Riemann approximation ver, it is $O(n^{-1/2})_{1}^{5-1}$ see that its speed of convergence is of the same order, regard support of f. T is not the $rac{1}{2}$ for other (deterministic) numerical integr he error e by the Riemann approximation using r_{i} imensional functi $n^{-1/2}$). 5 $|f(x) - f(\xi_{\text{mid}})| < \frac{\Delta}{2} \cdot \max |f'(x)| \text{ for } |x - \xi_{\text{mid}}| \le \frac{\Delta}{2}$ $|f(x)| - f(\xi_{\mathrm{mid}})| < \frac{\Delta}{2} \cdot \max |f'(x)| \text{ for } |x - \xi_{\mathrm{mid}}| \le \frac{\Delta}{2}$ 1 - $|f(x) - f(\xi_{\text{mid}})| < \frac{\Delta}{2} \cdot \max |f'(x)| < \frac{\Delta}{2} \cdot \max |f'(x$ 0 $\xi_{\rm mi}$

Computation of numerical integration by Rief

. 1.4. Illustration of numerical integration by Riemann sums $f_{\text{methods}} = f_{\text{methods}} = f_{\text{methods}} = \xi_{\text{methods}} = \xi_{\text{method$ This makes the Monte Cerle methods especting stratige to being relatively simple interra Monte Carlo method offers the advantage of being relatively simple and thus easy to implement of puter.

 $\widehat{I}_{n}^{(2)} = \frac{1}{2} \sum_{n=1}^{m-1} \int \left(\frac{1}{1+\frac{1}{2}} \frac{m_{n}}{n} + \frac{1}{2} \frac{m_{n}}{n} \right) \int \frac{1}{2} \int \frac{m_{n}}{n} \int \frac{1}{2} \int \frac{m_{n}}{n} \int \frac{1}{2} \int \frac{m_{n}}{n} \int \frac{m_{n}}{n} \int \frac{1}{2} \int \frac{m_{n}}{n} \int \frac$ 1.3 A Brief History of Monte Carlo Methodscall that, for out Monte provinger to gride is on fidelated interval was shrinking Experimental Mathematics is an old discipline: the Old Testament (1 Kings vii weimittis caster ascess habits is prove of the Ond Testament (1 Kings vii 2) contains a rough estimate of π (using the columns of King Solomon's temple). Monte Carlo method contains a rough estimate of π (using the columns of King Solomon's temple). Monte Carlo methesupport oare a somewhat more recent discipline. One of the first documented Monte Carlo experiments is *Buff* imensional daehillegiation (here the first) documented Monte Carlo experiments is *Buff* intensional daehillegiation (here the first) documented that this experiment can be use (1812) suggested that this experiment can be use (-1/2) 5 approximate π .

mple 1.3 (Buffon's needle). In 1733, the Grand of Buffon's needle) Experiment about the follow for the follow f bability that that a needle of f the stand on the floor will intersect one bability that a needle of f the stand of the stand on the floor will intersect one of the stand of the stan fon answered Buffon answered the quest in himself 14-1777 (Buffon, 1777)? Call that, for out Monte Carlo method the confidence interval was shrinking ume the needle failued sich ender its and cals of the figure \$.5). Then the fue fue the the the the the ver, it is easy to see that its speed of donvergence is of the same order regard

Interpretation

- Say found the first two decimal for the integral 0.45??? using a naive numerical integration...
 - in Id, to get one more decimal correct, need I0x more work
 - in 2d, to get one more decimal correct, need 100x more work
 - in 3d, to get one more decimal correct, need 1000x more work



Exerc. 10

Enters Simple Monte Carlo...



- Use this plot to empirically derive the running time of Simple Monte Carlo for a given tolerence *tol*
- Create the plot for Example 7

Exercise: