

2

2.0

Review / Q&A

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 $\widehat{I}_n = \frac{1}{n} \sum_{i=1}^n \phi(X_i)$.

Optimization and simulation

x : unknown y : data

Cases where this is advantageous...

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



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



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

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 (2 digits: I00x)
 - in 2d, to get one more decimal correct, need 100x more work (2 digits: 10,000x)
 - in 3d, to get one more decimal correct, need 1000x more work (2 digits: 1e6x)



Exerc. 10 Enters Simple Monte Carlo...



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

Thm. 12

Theoretical foundations

- Law of large number for iid random variable
- Central limit theorem for iid random variable
- IMPORTANT: there are LLNs and CLTs for dependent random variable (later!)

Exact sampling methods

Exact sampling: simple cases

- Inversion method
- Transformation method
- Augmentation method

Inversion method

Def. 13

 Consider a real-valued random variable X and its associated cumulative distribution function (cdf)

$$F(x) = \mathbb{P}(X \le x) = F(x)$$

• The cdf $F : \mathbb{R} \to [0, 1]$ is

- increasing; i.e. if $x \leq y$ then $F(x) \leq F(y)$
- right continuous; i.e. $F(x + \epsilon) \rightarrow F(x)$ as $\epsilon \rightarrow 0$ ($\epsilon > 0$)
- $F(x) \rightarrow 0$ as $x \rightarrow -\infty$ and $F(x) \rightarrow 1$ as $x \rightarrow +\infty$.
- We define the generalised inverse

$$F^{-}(u) = \inf \{x \in \mathbb{R}; F(x) \ge u\}$$

also known as the quantile function. Note that $F^{-}(u) = F^{-1}(u)$ if F is continuous.



• **Proposition**. Let *F* be a cdf and $U \sim U_{[0,1]}$. Then $X = F^{-}(U)$ has cdf *F*.

Proof. $F^{-}(u) \leq x \Leftrightarrow u \leq F(x)$ so $U \sim \mathcal{U}_{[0,1]}$, we have

$$\mathbb{P}\left(F^{-}\left(U\right) \leq x\right) = \mathbb{P}\left(U \leq F\left(x\right)\right) = F\left(x\right).$$

Exercise: construct a RNG for exponential random variables of a given rate

What about multivariate distributions?

Transformation method

- Let $Y \sim q$ be a \mathbb{Y} -valued random variable (rv) of, which we can simulate (eg, by inversion)
- Let $X \sim \pi$ be a X-valued rv, which we wish to simulate.
- It may be that we can find a function φ : Y → X with the property that if we simulate Y ~ q and then set X = φ(Y) then we get X ~ π.
- Inversion is a special case of this idea.

Augmentation / auxiliary variables Def. 16

- Example: how to simulate from a mixture model?
- Key idea: marginalization is easy with Monte Carlo methods
 - Contrast with analytical marginalization
- Build (X,Y) such that distribution of interest is a marginal
- Exercises:
 - write pseudo-code to simulate from a mixture distribution with 2 normal components,
 - show that rejection sampling is an augmentation sampling scheme

Running time of Monte Carlo methods

Fundamental equation to analyze Monte Carlo methods

Х

running time =

number of samples needed to get a tolerance (with probability 95%)

compute cost per sample

- Exercise:
 - Compute the running time in tol and d for Example 7 but with non-diagonal covariance normal vectors

Using CLT in practice

- Let X be a random tree
- We are looking at a clade indicator f(X) as in Example 2.
- After 500 iid trees your MC estimate for the clade support is roughly 10%
- Should you extract more samples?
 - Say we want a scheme with relative error of less than 10% for approximately 95% of the random seeds