

simulation-based approximation of posterior distributions

STAT 206  
25 Feb 21

Monte Carlo methods

history of Monte Carlo  
①

(1730) Buffon needle (physical)

(1908) Mr. Garret (physical)

t. story

(~1940) Bletchley Park project (England)

Manhattan project (US)

1st computers

Turing  
German  
origins  
Klopp

atom bomb (fission)

Feynman  
von Neumann  
Metropolis  
Ulam ...

How to usefully summarize (high-dimensional) probability distributions?  
↓ low-dimensional

$$\theta \sim (\theta_1, \dots, \theta_k)$$

D dataset

B

$$p(\theta \mid D [SM] [PM] B) =$$

$$c \cdot p(\theta \mid [PM] B) \cdot \rho(\theta \mid D [SM] B)$$

$$p(\theta_1 \mid D [SM] [PM] B)$$

$$(marginal) = \int \dots \int_{\leftarrow (k-1) \rightarrow} p(\theta \mid D [SM] [PM] B) d\theta_2 d\theta_3 \dots d\theta_k$$

$$E(\theta_1 | D[S_M][P_M] \mathcal{B}) =$$

(2)

$$\int \theta_1 p(\theta_1 | D[S_M][P_M] \mathcal{B}) d\theta_1$$

$$SD(\theta_1 | D[S_M][P_M] \mathcal{B}) =$$

$$\sqrt{V(\theta_1 | \dots)} = \sqrt{\frac{E(\theta_1^2 | \dots) - [E(\theta_1 | \dots)]^2}{}}$$

$$p(D^* | \dots) = \int p(D^* | \theta \sim) p(\theta | \dots) d\theta$$

Monte Carlo method:

Metropolis  
& Ulam  
~ 1942,  
pub. (1949)

Anything you want to know about  $p(\theta_1, \dots, \theta_k | \dots)$ , as long as  $k \geq 1$  is finite,

can be learned to arbitrary accuracy by making a (large) number of random draws from  $p(\theta_1, \dots, \theta_k | \dots)$

Q: How create algorithms, implementable <sup>(3)</sup> by digital computers, to make IID random draws from  $p(a_1, \dots, a_k | \sim)$ !

A: we have to settle for pseudo-random draws, because digital computers are

deterministic

behave in an apparently random fashion

First such algorithms:

$$U_i \sim \text{Uniform}(0, 1)$$

number theory

- ① take a big integer (<sup>30</sup> digits, <sup>10<sup>4</sup></sup> long)
- ② square it
- ③ extract middle 30 digits
- ④ normalize to (0, 1)
- ⑤ repeat

computational methods for  $U(0, 1)$  generation

inverse CDF  
method  
(probability  
in terms of  
transform)

$X$  continuous  
has CDF  $F_X(x)$  (4)

$$I \sim \boxed{F_X(X)} \sim U(0, 1)$$

$$\Leftrightarrow F_X^{-1}(U(0, 1)) \sim X$$

von Neumann:  
rejection  
sampling

from 1940s to today,  
thousands of papers have  
been published on finding  
efficient random sampling algorithms  
(algorithms)

2 goals  
in Monte  
Carlo

1 our method must be valid  
(i.e., it must make random  
draws that really follow target dist.)

2 want high degree of Monte Carlo  
efficiency (small clock time to run)

ⓑ has a general-purpose optimization built-in function called optim, for maximizing or minimizing functions from  $\mathbb{R}^k$  to  $\mathbb{R}^1$  (R code)

Bayesian analysis of the t SM in the NB10 case study

Our Bayesian model would be

$$(\mu, \sigma^2 | \mathcal{B}) \sim p(\mu, \sigma^2 | \mathcal{B})$$

$$(I_i | \mu, \sigma^2, \mathcal{B}) \stackrel{IID}{\sim} \text{tr}(\mu, \sigma^2)$$

$$(i = 1, \dots, n)$$

there is no conjugate prior

in this SM; from context we want a LI prior, but conjugacy can't help to create one.

So we need a new computation method <sup>Bayesian</sup> in the year 2021:

③ possibilities: ① Laplace approximation

~~and~~ ② simulation-based numerical ①  
5.2  
approximation of posterior distributions

---

③ large-sample Gaussians approximations

---

if time permits I'll sketch the Laplace approximation story in a discussion section

---

③ is simple to state by Bernstein-von Mises, for large  $n$  and a LI prior

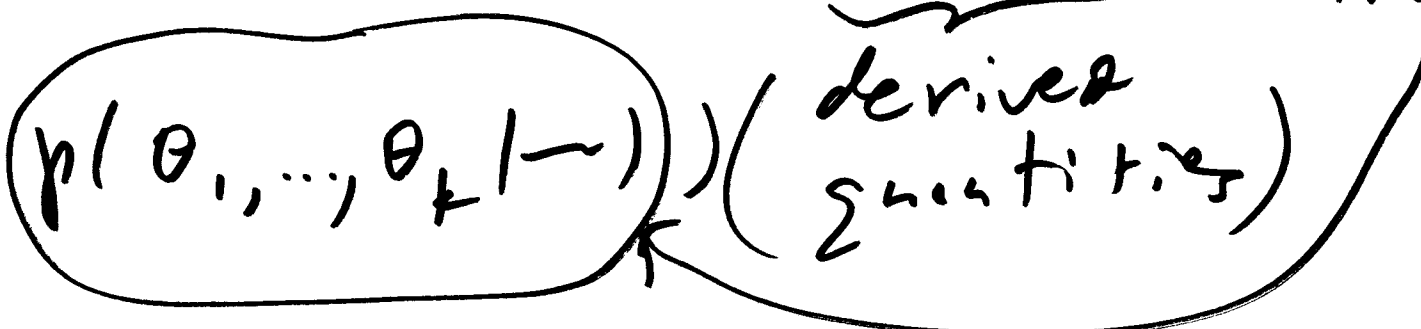
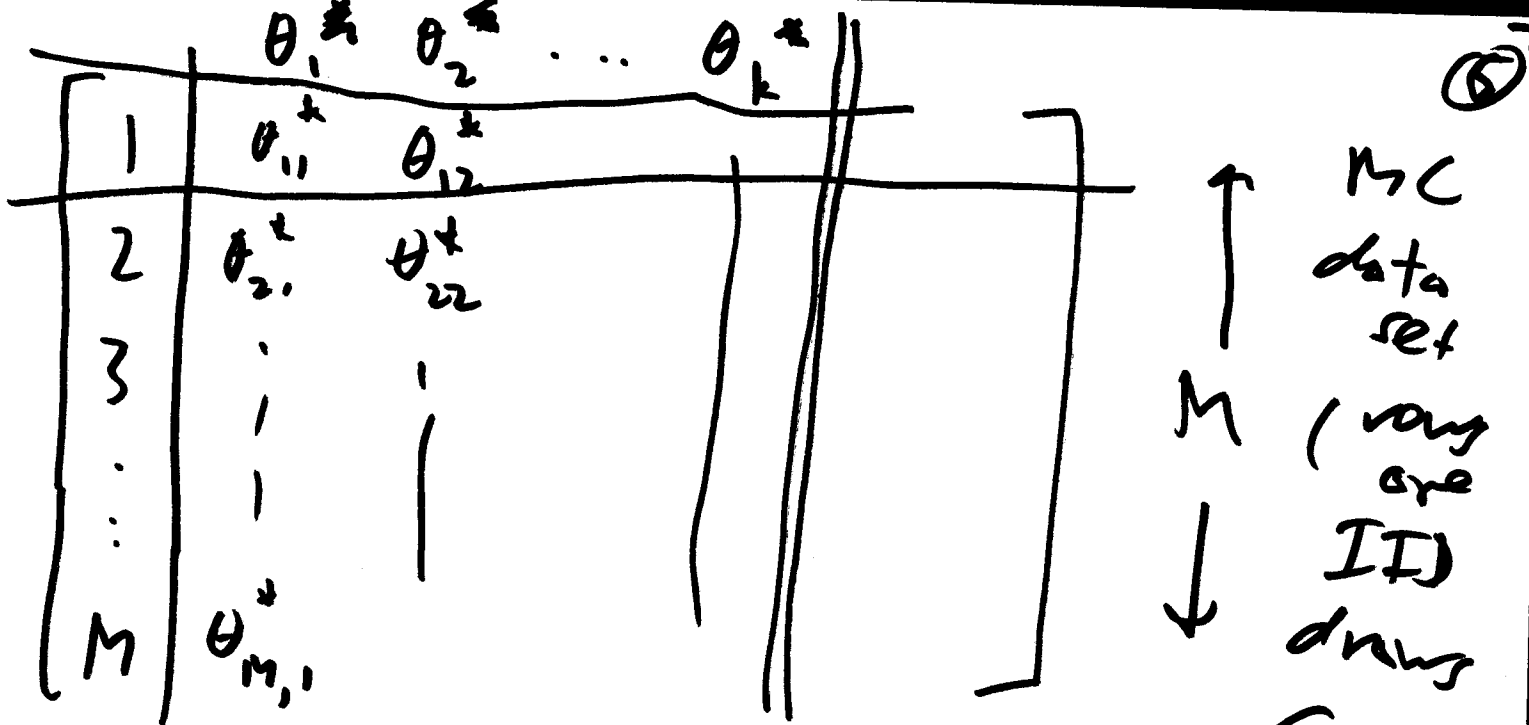
---

$$(\theta | D [SM] [PM] B) \sim N_k \left( \hat{\theta}_{MLE}^T, I^{-1} \right)$$

$\theta = (\theta_1, \dots, \theta_k)$

---

this produces exactly the same posterior (credible) intervals for  $(\theta_1, \dots, \theta_k)$  as with the large-sample maximum-likelihood approach's confidence intervals



Quiz 2  
 Case study  
 MC data set  
 $p(\theta) \propto [PM]B$

