simulation-based approximation of posterior distributions: go to (5.1)

STAT 206
25 Feb 21

Lecture

Monte Carlo methods

history of Monte Carlo (1)

(1730) Buffon needle (physical)

(1908) Mr. Gosset + t story (physical)

(~1940) Bletchley Park project (England) — Turing

German enigma decoding

1st computers

Manhattan project (US)

atom bomb (fission)

Feynman
von Neumann
Metropolis
Ulam ...

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

k-dimensional

$\underset{\sim}{\theta} = (\theta_1, \dots, \theta_k)$

$\mathcal{D}$ data set

$\mathcal{B}$

$$p(\underset{\sim}{\theta} \mid \mathcal{D}\,[SM]\,[PM]\,\mathcal{B}) =$$

$$c \cdot p(\underset{\sim}{\theta} \mid [PM]\,\mathcal{B}) \cdot \ell(\underset{\sim}{\theta} \mid \mathcal{D}\,[SM]\,\mathcal{B})$$

$p(\theta_1 \mid \mathcal{D}\,[SM]\,[PM]\,\mathcal{B})$

(marginal) $= \underbrace{\int \cdots \int}_{(k-1)} p(\underset{\sim}{\theta} \mid \mathcal{D}\,[SM]\,[PM]\,\mathcal{B})\, d\theta_2\, d\theta_3 \cdots d\theta_k$

$$E(\theta_1 \mid \mathcal{D}\,[SM]\,[PM]\,\mathcal{B}) =$$

$$\underline{\int \theta_1 \, p(\theta_1 \mid \mathcal{D}\,[SM]\,[PM]\,\mathcal{B})\, d\theta_1}$$

$$SD(\theta_1 \mid \mathcal{D}\,[SM]\,[PM]\,\mathcal{B}) =$$

$$\sqrt{V(\theta_1 \mid \cdots)} = \sqrt{\underline{E(\theta_1^2 \mid \cdots)} - [E(\theta_1 \mid \sim)]^2}$$

$$p(\mathcal{D}^* \mid \text{---}) = \int p(\mathcal{D}^* \mid \underset{\sim}{\theta} \sim)\, p(\underset{\sim}{\theta} \mid \text{---})\, d\underset{\sim}{\theta}$$

Monte Carlo (MC) method:

Metropolis & Ulam ~1942, pub. (1949)

Anything you want to know about $p(\theta_1, \ldots, \theta_k \mid \text{---})$, as long as $k \geq 1$ is finite, can be learned to arbitrary accuracy by making a (large) number of random draws from $p(\theta_1, \ldots, \theta_k \mid \text{---})$

Q:] How create algorithms, implementable by digital computers, to make IID random draws from $p(\theta_1, \dots, \theta_k \mid \sim)$?

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 (30 digits long)

(2) square it (3) extract middle 30 digits

(4) normalize to (0, 1)

(5) repeat

congruential methods for $U(0,1)$ generation inverse CDF method (probability integral transform)

$\underline{\underline{X}}$ continuous has CDF $F_X(x)$

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

$$\leftrightarrow \quad F_X^{-1}[\underline{U(0,1)}] \sim \underline{X}$$

von Neumann: rejection sampling

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

2 goals in Monte Carlo

[1] our method <u>must</u> be <u>valid</u> (i.e., it must make random draws that really follow target dist.)

[2] want high degree of Monte Carlo <u>efficiency</u> (small clock time to run)

(R) 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). (5.1)

Bayesian analysis of the t SM in the NB10 case study

Our Bayesian model would be

$$(\mu\ \sigma\ \nu \mid \mathcal{B}) \sim p(\mu\ \sigma\ \nu \mid \mathcal{B})$$
$$(y_i \mid \mu\ \sigma\ \nu\ \mathbb{T}\ \mathcal{B}) \stackrel{IID}{\sim} t_\nu(\mu, \sigma^2)$$
$$(i = 1, \dots, n)$$

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 Bayesian computation method in the year 2021: (3) possibilities: (1) Laplace approximation

② simulation-based numerical approximation of posterior distributions (6) (5.2)

---

③ large-sample Gaussian approximations

---

if time permits I'll sketch the (1) Laplace approximation story in a discussion section

---

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

$$(\underset{\sim}{\theta} \mid \mathcal{D}\,[SM]\,[PM]\,\mathcal{B}) \;\dot{\sim}\; N_k\left(\hat{\theta}_{MLE},\ \hat{I}^{-1}\right)$$

$\underset{\sim}{\theta} = (\theta_1, \dots, \theta_k)$

---

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

$$
\begin{array}{c|cccc}
 & \theta_1^* & \theta_2^* & \cdots & \theta_k^* \\
\hline
1 & \theta_{11}^* & \theta_{12}^* & & \\
2 & \theta_{21}^* & \theta_{22}^* & & \\
3 & \vdots & \vdots & & \\
\vdots & & & & \\
M & \theta_{M,1}^* & & &
\end{array}
\quad \uparrow M \downarrow
$$

MC data set (rows are IID draws from

$$p(\theta_1, \dots, \theta_k \mid \sim)$$

(derived quantities)

---

quiz 2 case study

MC data set

$$
\begin{array}{c|c}
 & \theta \\
\hline
1 & \theta_1^* \\
2 & \theta_2^* \\
\vdots & \vdots \\
M & \theta_M^*
\end{array}
\qquad M = 1{,}000
$$

$$p(\theta \mid y\, [PM]\, B)$$

$$
\begin{bmatrix}
0.915 & 1 \\
0.910 & 1 \\
0.884 & 0 \\
\vdots & \vdots \\
\vdots & \vdots
\end{bmatrix}
$$