STAT 425: Introduction to Bayesian Analysis

Marina Vannucci

Rice University, USA
Lecture 3: Monte Carlo simulation

- MC simulation methods
Monte Carlo Simulation

We have seen that in Bayesian Analysis statistical inferences are based on

$$p(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\int p(x|\theta)\pi(\theta)d\theta}$$

- This require to compute a complex integral at the denominator:

$$\int p(x|\theta)\pi(\theta)d\theta$$

- Conjugate case (e.g., normal-normal, beta-binomial) → easy to integrate things out and compute the quantities we need.
Monte Carlo Simulation

We have also seen that Bayesian use posterior summary statistics to summarize some information from the posterior distribution, for example the posterior mean:

\[ E[\theta|x) = \int \theta p(\theta|x)d\theta \]

and posterior credible intervals, which are based on quantiles \( z \) such that, for example,

\[ \int_{-\infty}^{z} p(\theta|x)d\theta = 0.025 \quad \text{or} \quad \int_{z}^{+\infty} p(\theta|x)d\theta = 0.975, \]

for the 95% credible interval.

- Once again, this is very easy to do in the conjugate case: the posterior is of the same distributional form as the prior, with the posterior parameters being simple updates of the prior parameters.
Monte Carlo Simulation

- But what if we cannot compute those integrals in closed form?
- This might happen for example if the model is not conjugate:
  - “Non-standard” likelihood
  - more complex models (e.g., multi-priors for different parameters)
  - Prior info captured better by non-conjugate prior distributions (e.g. a truncated normal instead of a simple normal for a normal mean, or a normal truncated on \([0, 1]\) for the probability of success)
- Some other times, the predictive distribution may not be computable in closed form, as a complex integral

\[
p(x_{n+1}|x) = \int p(x_{n+1}|\theta) p(\theta|x) \, d(\theta)
\]

- In all the cases where the pen fails, we can use Monte Carlo simulation to obtain information on the posterior distribution.
Ordinary Monte Carlo Simulation (OMC) –1

- The “Monte Carlo method” refers to the theory and practice of learning about probability distributions by simulation rather than calculus.

- In ordinary Monte Carlo (OMC) we use IID simulations from the distribution of interest: suppose \( \theta_1, \theta_2, \ldots, \theta_m \) are IID samples from some distribution, and suppose we want to know an expectation

\[
\mu = E\{g(\theta)\} = \int g(\theta)p(\theta)d\theta
\]

which is the expectation of a function, \( g(\cdot) \), of a random variable, \( \theta \), having density \( p(\cdot) \). The law of large numbers (LLN) says that if we can obtain a sample of \( m \) values from \( p, \theta_i \overset{i.i.d.}{\sim} p(\cdot) \), then

\[
\hat{\mu}_n = \frac{1}{m} \sum_{i=1}^{n} g(\theta_i)
\]

converges in probability to \( \mu \).
Ordinary Monte Carlo Simulation (OMC) – 2

Similarly for the standard deviations,

\[ \hat{\sigma}_m = \sqrt{\frac{1}{m} \sum_{i=1}^{m} [g(\theta_i) - \hat{\mu}_n]^2} \rightarrow sd(\theta), \]

and for the quantiles, etc....

In a OMC:
- the “data” \( \theta_1, \ldots, \theta_m \) are computer simulations rather than measurements on objects in the real world,
- the “sample size” is the number of computer simulations rather than the size of some real world data, and
- the unknown parameter \( \mu \) is in principle completely known, given by some integral, which, unfortunately, we are unable to do.
OMC Terminology

- Often we say that $m$ is the *Monte Carlo sample size* to distinguish it from anything else that may be called a sample size.
- Often we say that $\hat{\mu}_m$ is the *Monte Carlo estimate* or *Monte Carlo approximation* or *Monte Carlo calculation* of $\mu$ to distinguish it from anything else that may be called an estimate.
- Often we say that $\hat{\sigma}_m / \sqrt{m}$ is the *Monte Carlo standard error* of $\hat{\mu}_n$ to distinguish it from anything else that may be called a standard error.
Other examples

- Cumulative ordered values approximate $F(\theta)$ (cdf)
- Empirical distribution of the sample $\theta_1, ..., \theta_m$ approximates $p(\theta)$ (use histogram or kernel density estimator)
- $P(g(\theta) > c)$ approximated by proportion of samples where event $(g(\theta^{(i)}) > c)$ occurs
- Sample moments/quantiles/functions approximate true moments/quantiles/functions

Extends easily to higher dimensional parameters.
Normal quantiles

Suppose that $\theta \sim N(0, 1)$ and we are interested in

$$Pr(\theta \geq 1) = \int_{1}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}} d\theta.$$

We cannot compute this integral in closed form

Monte Carlo Integration:

```r
n=100000
theta=rnorm(n, 0,1)  ## n samples from a N(0,1)
mean_theta=mean(theta)  ## expected value E(\theta)
sd_theta=sd(theta)  ## sd(\theta)
Pr_theta_geq_1=mean(theta>=1)  ## true value approx 0.16
mean_theta; sd_theta; Pr_theta_geq_1;
```

```
[1] 0.006006067
[1] 0.995561
[1] 0.15784
```
Monte Carlo variance

- If
  \[ \int \theta^2 p(\theta) d\theta < +\infty \]

  the central limit theorem shows that

  \[ \frac{\bar{\theta}_{MC} - \int \theta p(\theta) d\theta}{\sqrt{\sigma^2/m}} \rightarrow N(0, 1) \]

  where \( \sigma^2 = \text{Var}[\theta] \). We can obtain the Monte Carlo standard error by means of

  \[ \text{Var}[\theta_{MC}] \approx \hat{\sigma}^2 = \frac{1}{m} \sum [\theta^{(i)} - \bar{\theta}_{MC}]^2 \]

- An approximate 95% Monte Carlo confidence interval for the mean of \( \theta \) is then \( \bar{\theta}_{MC} \pm 1.96 \sqrt{\hat{\sigma}^2/m} \), that is we expect the posterior mean of \( \theta \) to be in this interval for roughly 95% of repeated MC samples.

- To increase accuracy, increase \( m \).
Monte Carlo variance

- Similarly for $g(\theta)$, if

$$\int [g(\theta)]^2 p(\theta) d\theta < +\infty$$

the central limit theorem shows that

$$\frac{\overline{g(\theta)}_{\text{MC}} - \int g(\theta) p(\theta) d\theta}{\sqrt{\sigma_g^2/m}} \rightarrow N(0, 1)$$

where $\sigma_g^2 = \text{Var}[g(\theta)]$. We can obtain confidence bounds for $\overline{g(\theta)}_{\text{MC}}$ and estimate the asymptotic variance by means of

$$\text{Var}[g(\theta)_{\text{MC}}] \simeq \hat{\sigma}_g^2 = \frac{1}{m} \sum [g(\theta^{(i)}) - \overline{g(\theta)}_{\text{MC}}]^2$$

- An approximate 95% Monte Carlo confidence interval for the mean of $g(\theta)$ is $\overline{g(\theta)}_{\text{MC}} \pm 1.96 \sqrt{\hat{\sigma}_g^2/m}$. 