Lecture 8: Bayesian Estimation: Direct Sampling

Big Data and Machine Learning for Applied Economics
Econ 4676

Ignacio Sarmiento-Barbieri

Universidad de los Andes

September 2, 2021

Agenda

– 16 de sept

1. Bayesian Estimation
2. Simulation-based methods for Bayesian analysis
   - Direct Sampling

   – Gibbs Sampling
3. Recap
4. Further Readings

Bayesian Estimation

$\underbrace{X_1, \quad X_n}$ $\sim \mathcal{N}(\mu, 1)$

$\hat{\mu} = \bar{X}$

- The Bayesian approach to stats is fundamentally different from the classical approach we have been taking
- In the classical approach, the parameter β is thought to be an unknown, but fixed quantity, e.g., $X_i \sim f(\beta)$
- In the Bayesian approach β is considered to be a quantity whose variation can be described by a probability distribution (*prior distribution*)
- Then a sample is taken from a population indexed by β and the prior is updated with this sample
- The resulting updated prior is the *posterior distribution*

Bayes Approach

Bayes Theorem

$$\underbrace{\pi(\beta|X)}_{\text{posterior}} = \frac{\overbrace{f(X|\beta)}^{\text{densidad}}\,\overbrace{p(\beta)}^{\text{prior}}}{\underbrace{m(X)}_{\text{marginal}}} \quad \Rightarrow \quad \underbrace{f(X,\beta)}\, = f(X|\beta)\,p(\beta) \tag{1}$$

with $m(X)$ is the marginal distribution of X, i.e.

$$m(X) = \underbrace{\int f(X|\beta)p(\beta)d\beta} \tag{2}$$

It is important to note that Bayes' theorem does not tell us what our beliefs should be, it tells us how they should change after seeing new information.

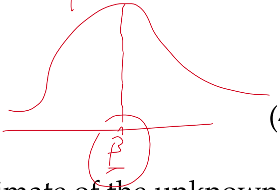
Frequentist Approach

- The interest is on β, frequentist estimation procedures give us that, for example

$$\hat{\beta}_{MLE} = (X'X)^{-1}X'y \tag{3}$$

$\beta \overset{Bayes}{\sim} \pi(\beta|X)$

- Now in Bayes world, I have the full distribution. Which β I use?
- I can use any moment, but usually the interest lies on



$$E(\beta) = \int \beta\pi(\beta|X) \tag{4}$$

$(\theta - \hat{\theta})^2$

- Why? note that if you use MSE as loss function, the Bayes estimate of the unknown parameter is the mean of the posterior distribution

$E(\beta_{posterior}) = \omega \beta_{MLE} + (1-\omega)\beta_0$

CEF

↳ HW

Bayesian Estimation

- We are going to have an overview simulation-based methods for Bayesian analysis.
  1. Direct sampling algorithm
  2. Gibbs sampling algorithm
- As a running example, we use the linear regression framework

$$y_i = \beta x_i + u_i\ , \quad u_i \sim N(0, \sigma^2) \tag{5}$$

no constant

- with σ^2 known, and
- with prior distribution $\beta \sim N(\beta_0, \tau^2)$

σ^2 unknown $\quad \sigma^2 \sim f(\cdot)$

Direct Sampling

$A \quad , \quad B$

$m = -\frac{1}{2}\frac{B}{A}$

$V = \frac{1}{A}$

- Using the knowledge of conjugate priors + the trick for exponentials
- The posterior distribution β follows the normal distribution:

$$\beta | Y, X \sim N\left(\underbrace{\frac{\frac{1}{\sigma^2}\sum_{i=1}^{N} y_i x_i + \frac{1}{\tau^2}\beta_0}{\frac{1}{\sigma^2}\sum_{i=1}^{N} x_i^2 + \frac{1}{\tau^2}}}_{m}, \underbrace{\frac{1}{\frac{1}{\sigma^2}\sum_{i=1}^{N} x_i^2 + \frac{1}{\tau^2}}}_{Var} \right) \tag{6}$$

- We where able to characterize the full posterior distribution for the unknown object.

Direct Sampling

- Suppose now, that our object of interest is not β per se, but some nonlinear function of unknown parameter β, e.g. $h(\beta)$.
- For example:
  - $h(\beta) = \beta$
  - $h(\beta) = |\beta|$ ✓
  - $h(\beta) = \alpha\%$ *quantile of* β ✓
  - $h(\beta) = \beta^3$ ✓
  - $h(\beta) = \beta_1\beta_2$ ✓
- The goal is to obtain posterior moments of $h(\beta)$.

$$y = \beta_1 Room + \beta_2 Room^2$$

$$\beta_1 + 2\beta_2 Room = 0$$

$$\boxed{Room = -\frac{\beta_1}{2\beta_2}}$$

Direct Sampling

Side note: Frequentist´s approach

- Frequentist obtain the sampling distribution of $h(\beta)$ using the delta method:
- If we have CLT

$$\underbrace{\sqrt{N}\left(\widehat{\beta}-\beta_0\right)} \rightarrow_d \mathrm{N}\left(0, V_{asy}\right) \tag{7}$$

- Then we

$$\sqrt{N}\left(h\left(\widehat{\beta}\right)-h(\beta_0)\right) \rightarrow_d \mathrm{N}\left(0, \underbrace{V_{asy}\left[h'(\beta_0)\right]^2}\right) \tag{8}$$

- As $N \rightarrow \infty$ where ~~n~~ N is the number of observations.

Direct Sampling

$\pi(\beta|X) \sim \underline{\underline{N}}(m, V)$

- Idea: Monte Carlo integration.
- Requirement
  - Know how to generate i.i.d. samples from the posterior distribution of β, $\pi(\beta|Y)$
- The requirement is satisfied for our linear regression example:
  - The posterior distribution of β follows the normal distribution.
  - Most modern statistical program languages provide random number generators for many parametric distributions including the normal distribution.

set seed (1001)

set seed

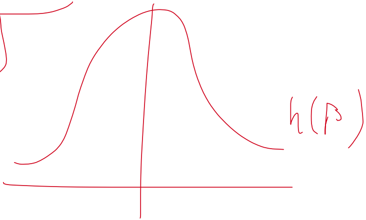
seed

Direct Sampling

- Direct sampling approach simply approximates the posterior expectations of a function $h(\beta)$ by

$$E_Y^{\beta}\left[h\left(\beta\right)\right] = \int h\left(\beta\right)\pi\left(\beta|Y\right)d\beta \tag{9}$$

$$\approx \frac{1}{S}\sum_{i=1}^{S} h(\beta^i)$$



- Where β^i is i.i.d. samples from $\pi(\beta|Y)$
- S is "number of random samples from the posterior" or "number of generated draws" NOT the number of observations.

No confundir con Bootstrap

Direct Sampling

- Provided that $E_Y^\beta\left[h(\beta)^2\right] < \infty$,
- we can use the Strong Law of Large Numbers (SLLN)

$$\frac{1}{S}\sum_{i=1}^{S} h(\beta^i) \rightarrow a.s. \int h(\beta)\, p(\beta|Y)\, d\beta$$

- and the Central Limit Theorem (CLT)

$$\sqrt{S}\left(\frac{1}{S}\sum_{i=1}^{N} h(\beta^i) - \int h(\beta)\, \pi(\beta|Y)\, d\beta\right) \rightarrow_d N(0, V_\pi)$$

caracterizar
cuán precisa
es nuestra
aprox

- Where

$$V_\pi = \mathrm{Var}_Y^\beta\left(h(\beta)\right) = \int \left(h(\beta) - E_Y^\beta\left[h(\beta)\right]\right)^2 \pi(\beta|Y)\, d\beta$$

- S is the number of simulated draws from the posterior distribution

Direct Sampling

- Note that we turned a complicated integration into a simple average

$$\frac{1}{S}\sum_{i=1}^{S} h(\beta^i) \rightarrow a.s. \int h(\beta)\, p(\beta|Y)\, d\beta \tag{10}$$

- As the number of simulated draws increases, this simple average converges to the object of interest.
- Numerical accuracy?

Direct Sampling

- The CLT result provides a way to measure the numerical accuracy of this
- Monte Carlo approximation:

$$\sqrt{S}\left(\frac{1}{S}\sum_{i=1}^{N}h(\beta^i)-\int h\left(\beta\right)p\left(\beta|Y\right)d\beta\right)\rightarrow_d\ N(0,\ V_\pi) \tag{11}$$

- That is,

H W

$$\frac{1}{S}\sum_{i=1}^{S}h(\beta^i)\approx_d N\left(E_Y^\beta\left[h\left(\beta\right)\right],\frac{V_\pi}{S}\right) \tag{12}$$

- Where $V_\pi=\ \mathrm{Var}_Y^\beta\left(h\left(\beta\right)\right)$. Posterior variance of $h\left(\beta\right)$ scaled by $1/S$ determines the numerical accuracy. As $S\rightarrow\infty$, numerical approximation goes to zero
- Trade-off
  - Large $\cancel{N}$ S: high computational cost (time) but more accurate approximation
  - Small $\cancel{N}$ S: low computation cost (time) but less accurate approximation

Direct Sampling

Example: Linear regression

- Consider the following linear regression model

$$y_i = \beta x_i + u_i, \quad \mathrm{u_i} \sim N\left(0, \sigma^2\right) \tag{13}$$

- with prior distribution$\beta \sim N\left(\beta_0, \tau^2\right)$, and suppose σ^2 is known.
- Then, we now all know that the posterior distribution β follows the normal distribution:

$$\beta|Y,X \sim \underbrace{N\left(\frac{\frac{1}{\sigma^2}\sum_{i=1}^N y_i x_i + \frac{1}{\tau^2}\beta_0}{\frac{1}{\sigma^2}\sum_{i=1}^N x_i^2 + \frac{1}{\tau^2}}, \frac{1}{\frac{1}{\sigma^2}\sum_{i=1}^N x_i^2 + \frac{1}{\tau^2}}\right)}_{\text{POSTERIOR}} \tag{14}$$

$h(\beta) = |\beta|$

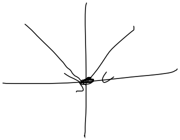
Direct Sampling

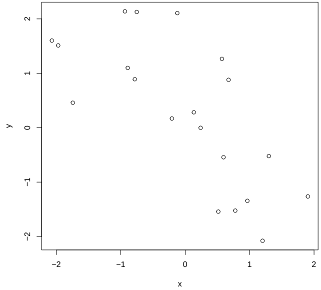
Example: Linear regression

- Goal: posterior mean and equal-tail-probability credible set for $|\beta|$ → valor absoluto
- I generate data (y_i, x_i) with
  - $y_i = \beta x_i + u_i, \quad u_i \sim N(0, \sigma^2)$
  - $N = 20$
  - $\boxed{\beta_{\text{true}} = -1}$ and $\sigma^2 = 1$
  - $\beta_0 = 0$ and $t = 100$

↳ 100

$|\beta_{true}| = 1$





CLT

medias ~ $N(m, V)$

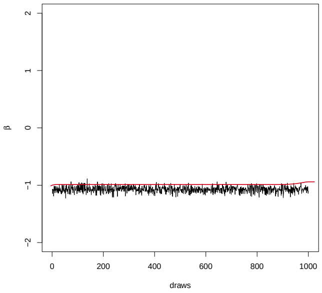
$m(\theta)$

Direct Sampling

Example: Linear regression

- Step 1: we generate N draws from the $N(m, V)$, $\{\beta^i\}\ 1, \ldots, S$
- $m = -1.07$
- $V = 0.0510$

Figure 1: Example of draws $\left(\{\beta^i\}_{1,\ldots,N}\right)$, $S = 1,000$

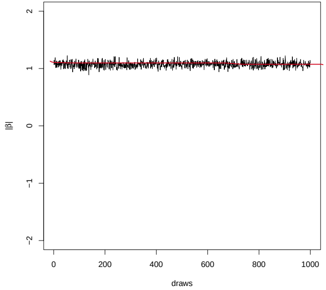


Direct Sampling

Example: Linear regression

- Step 2: we are interested in posterior moments of $|\beta|$.
- Turn draws into $\{|\beta|\}_{1,\ldots,S}$

Figure 2: Example of draws $\left(\{|\beta^i|\}_{1,\ldots,S}\right)$

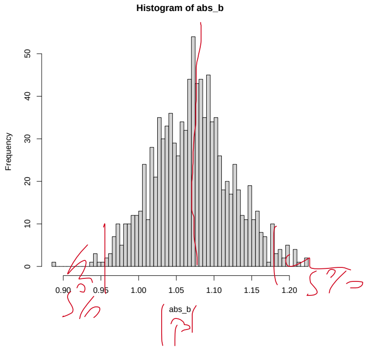


Direct Sampling

Example: Linear regression

- Histogram approximation to $\pi\left(|\beta| \;|Y\right)$ using $\left\{|\beta^i|\right\}_{1,\ldots,S}$

Figure 3: Example of draws $\left(\left\{|\beta^i|\right\}_{1,\ldots,S}\right)$



Direct Sampling

Example: Linear regression

- The posterior mean of $|\beta|$ is approximated by

$$E_Y^{\beta}\left[|\beta|\right] \approx \frac{1}{S}\sum_{i=1}^{S}\left|\beta^i\right| = 1.0719 \tag{15}$$

- The 90% equal-tail-probability interval is approximated by

$$C_Y = [q_1, q_u] = [0.719, 1.441] \tag{16}$$

- Where q_1 and q_u such that

$$5\% = \frac{1}{S}\sum_{i=1}^{S} 1\{|\beta| < q_u\} \qquad = 0{,}05 \tag{17}$$

Direct Sampling

Example: Linear regression

- Numerical accuracy of $\frac{1}{\cancel{N}\,S} \sum_{i=1}^{\cancel{N}\,S} 1 \left|\beta^i\right|$
- We know that if we generate enough number of β^i, we get an accurate approximation to the posterior moments
- How many draws are enough?
- In other words, "Will I get different answer if I construct the same quantity using different set of draws $\{\beta^i\}$?
- Is $S = 10$ enough? Or, is $S = 10,000$ enough?

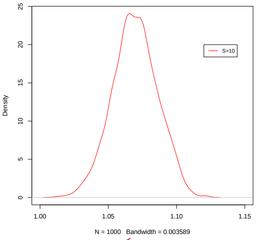
Direct Sampling

Example: Linear regression

$\frac{1}{S}\sum|\beta^i|$

- To see the numerical error I generate 1,000 sets of $\{\beta^i\}_{i=1,\ldots,S}$
- Compute 1,000 of $\frac{1}{S}\sum_{i=1}^{S}|\beta^i|$ to see how variable this Monte Carlo approximation with different S

Figure 4: Distribution of $\frac{1}{S}\sum_{i=1}^{S}|\beta^i|$ over $\{\beta^i\}_{i=1,\ldots,S}$



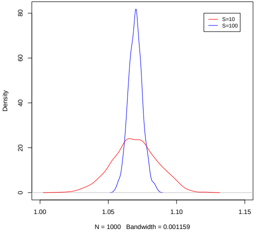
$\mathrm{SD}\ \frac{1}{\cancel{N}\ S}\sum_{i=1}^{\cancel{N}\ S}|\beta^i| = 0.688$

Direct Sampling

Example: Linear regression

- To see the numerical error I generate 1,000 sets of $\{\beta^i\}_{i=1,\ldots,S}$
- Compute 1,000 of $\frac{1}{S}\sum_{i=1}^{S}\left|\beta^i\right|$ to see how variable this Monte Carlo approximation with different S

Figure 5: Distribution of $\frac{1}{S}\sum_{i=1}^{S}\left|\beta^i\right|$ over $\{\beta^i\}_{i=1,\ldots,S}$



M = 1,07

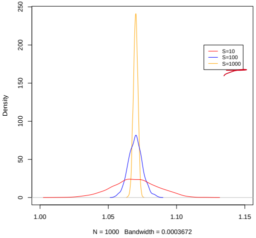
$$\text{SD}\ \frac{1}{\cancel{N}_{S}} \sum_{i=1}^{\cancel{N}\ S} \left|\beta^i\right| = \underline{0.022}$$

Direct Sampling

Example: Linear regression

- To see the numerical error I generate 1,000 sets of $\{\beta^i\}_{i=1,\ldots,S}$
- Compute 1,000 of $\frac{1}{S}\sum_{i=1}^{S}|\beta^i|$ to see how variable this Monte Carlo approximation with different S

Figure 6: Distribution of $\frac{1}{S}\sum_{i=1}^{S}|\beta^i|$ over $\{\beta^i\}_{i=1,\ldots,S}$



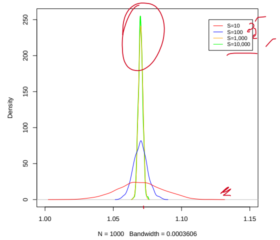
$$\mathrm{SD}\ \frac{1}{\cancel{N}\ S}\sum_{i=1}^{\cancel{N}\ S}\left|\beta^i\right| = 0.0074$$

Direct Sampling

Example: Linear regression

- To see the numerical error I generate 1,000 sets of $\{\beta^i\}_{i=1,\ldots,S}$
- Compute 1,000 of $\frac{1}{S}\sum_{i=1}^{S}|\beta^i|$ to see how variable this Monte Carlo approximation with different S

Figure 7: Distribution of $\frac{1}{S}\sum_{i=1}^{S}|\beta^i|$ over $\{\beta^i\}_{i=1,\ldots,S}$



$$\text{SD}\ \frac{1}{\underset{S}{\cancel{N}}}\sum_{i=1}^{\overset{S}{\cancel{M}}}|\beta^i| = 0.0023$$

Direct Sampling

Example: Linear regression

- What do we try to capture in this exercise?
- We try to mimic the distribution of Monte Carlo approximation offered by

CLT

$$\underbrace{\sqrt{S}\left(\frac{1}{S}\sum_{i=1}^{N} h\left(\beta^{i}\right) - \int h\left(\beta\right)\pi\left(\beta|Y\right)d\beta\right) \to_d N(0, V_\pi)}_{} \tag{18}$$

- All variation in this Monte Carlo approximation is due to numerical "simulation".
- Throughout this example, we fix $Y_{1:N}$, $X_{1:N}$ (not a sampling variation).

Recap

- If you know how to generate i.i.d draws from the posterior distribution of β,
- You also can posterior moments of $h(\beta)$ by simple average:

$$\int \text{h}\ (\beta)\, p\,(\beta|Y)\, d\beta \approx \frac{1}{\cancel{N}\ S} \sum_{i=1}^{\cancel{N}\ S} h(\beta^i) \tag{19}$$

- SLLN guarantees this Monte Carlo average to the right limit:

$$h = \frac{\beta_1}{2\beta_2}$$

$$\frac{1}{\cancel{N}\ S} \sum_{i=1}^{\cancel{N}\ S} h\left(\beta^i\right) \rightarrow_{\text{a.s}} \int h\,(\beta)\, \pi\,(\beta|Y)\, d\beta \tag{20}$$

- CLT tells you that the Monte Carlo average always has a numerical error:

$$\sqrt{S}\left(\frac{1}{S}\sum_{i=1}^{N} h\left(\beta^i\right) - \int h\,(\beta)\, \pi\,(\beta|Y)\, d\beta\right) \rightarrow_d N\,(0, V_\pi) \tag{21}$$

- It is important to check how good is your numerical approximation

Review & Next Steps

- Direct Sampler → iid → probabilidad/distribución
- **Next Class:** Gibbs Sampler

  ↳ secuencial

  → de cond distribución

  $y_i \rightarrow d_{i-1}$

Further Readings

- Casella, G., & Berger, R. L. (2002). Statistical inference (Vol. 2, pp. 337-472). Pacific Grove, CA: Duxbury. Chapter 7
- Hoff, P. D. (2009). A first course in Bayesian statistical methods (Vol. 580). New York: Springer.

→ Casella → Introduction to Monte Carlo

→ Applied Bayesian Statistics

Casella

Open BUGS *