Lab 4

February 5, 2013

HW 3

- Student's t distribution:

$$f(x\,|\,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{1}$$

  - Mean exists if $\nu > 1$.
  - Variance exists if $\nu > 2$.

- Cauchy distribution:

$$f(x\,|\,\nu = 1) = \frac{1}{\pi(1 + x^2)}. \tag{2}$$

Monte Carlo Simulation

- $\theta \in \Theta$: parameter.
- $y^n = (y_1, \ldots, y_n)'$: data.
- $g(\theta)$: function of θ.
- $p(\theta | y^n)$: posterior.
- Consider expectation of $g(\theta))$ with respect to the posterior,

$$E[g(\theta) | y^n] = \int_{\Theta} g(\theta) p(\theta | y^n) d\theta. \tag{3}$$

- How to compute the integral?

Law of Large Numbers

- For large S, generate a sample of S values from the posterior distribution,

$$\theta^{(1)}, \ldots, \theta^{(S)} \overset{\text{i.i.d.}}{\sim} p(\theta \mid y^n). \tag{4}$$

- LLN:

$$\frac{1}{S} \sum_{i=1}^{S} g\left(\theta^{(i)}\right) \rightarrow E\left[g(\theta) \mid y^n\right]. \tag{5}$$

- If $g(\theta) = \theta$,

$$\frac{1}{S} \sum_{i=1}^{S} \theta^{(i)} \rightarrow E\left[\theta \mid y^n\right]. \tag{6}$$

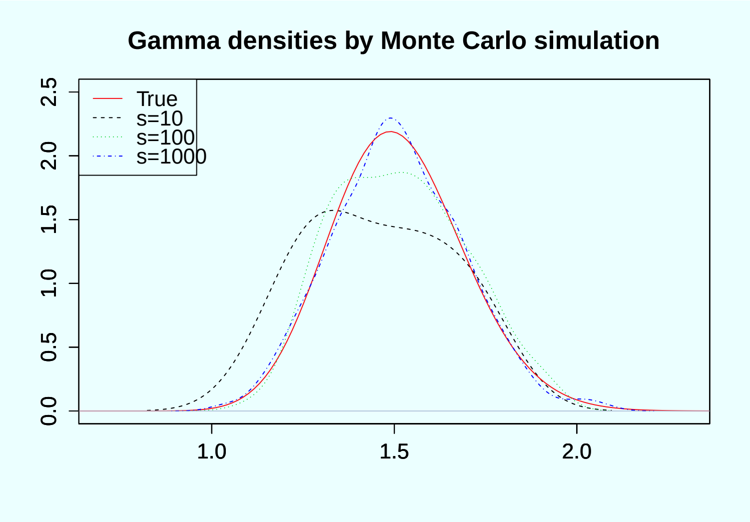
- If $g(\theta) = I_{(\theta > c)}$,

$$\frac{1}{S}\sum_{i=1}^{S} I_{(\theta^{(i)} > c)} \rightarrow E\left[I_{(\theta > c)}\,|\,y^n\right] = P[\theta > c\,|\,y^n]. \tag{7}$$

- The empirical distribution of $\left\{\theta^{(1)}, \ldots, \theta^{(S)}\right\} \rightarrow p(\theta\,|\,y^n)$
- Sample moments, quantiles and functions approximate true moments, quantiles and functions.

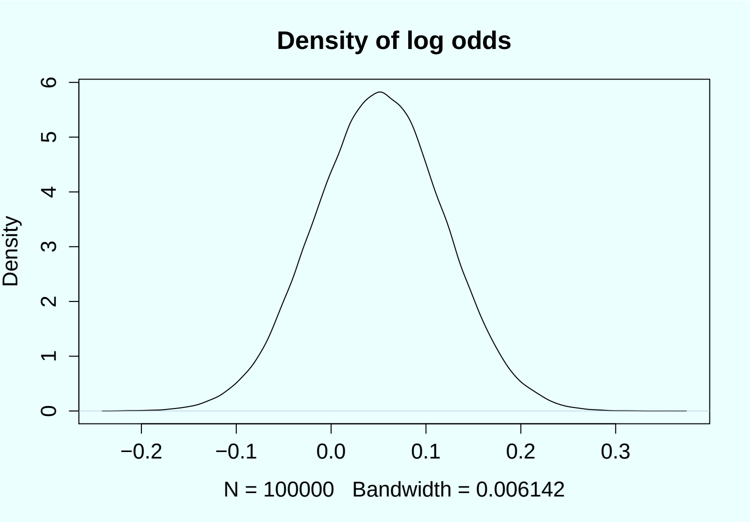
Example 1

- Assume the posterior $p(\theta \,|\, y^n)$ is Gamma(68,45).
- Generate $S = 10, 100, 1000$ samples from the posterior distribution and compute the empirical distributions.
- $P[\theta < 1.2 \,|\, y^n]$
  - True: 0.0367
  - S=10: 0.1
  - S=100: 0.01
  - S=1,000: 0.039


Gamma densities by Monte Carlo simulation
True
s=10
s=100
s=1000
0.0
0.5
1.0
1.5
2.0
2.5
1.0
1.5
2.0

Example 2

- θ: population proportion agree with some political policy.
- The posterior $p(\theta | y^n)$ is Beta(442,420).
- Consider $\gamma \equiv g(\theta) = \log \frac{\theta}{1-\theta}$.
- Empirical ratio is 0.51.
- We want to estimate $p(\gamma | y^n)$:
  1. Generate $\theta^{(i)} \sim p(\theta | y^n)$.
  2. Compute $\gamma^{(i)} = g(\theta^{(i)})$.
- The empirical distribution of $\{\gamma^{(1)}, \ldots, \gamma^{(S)}\} \to p(\gamma | y^n)$.


Density of log odds
Density
0
1
2
3
4
5
6
−0.2
−0.1
0.0
0.1
0.2
0.3
N = 100000 Bandwidth = 0.006142

Posterior Predictive Distribution

- $p(y \mid y^n)$: posterior predictive distribution.

$$p(y \mid y^n) = \int L(y \mid \theta) p(\theta \mid y^n) d\theta. \tag{8}$$

- $L(y \mid \theta)$: data generating function.
- $p(\theta \mid y^n)$: posterior distribution.
- Generate random samples from $p(y \mid y^n)$.
    1. Generate $\theta^{(i)} \sim p(\theta \mid y^n)$.
    2. Generate $y^{(i)} \sim L(y \mid \theta^{(i)})$.

- We can compute

$$E[g(y)] = \int g(y)p(y\,|\,y^n)dy, \tag{9}$$

$$= \frac{1}{S}\sum_{i=1}^{S} g(y^i). \tag{10}$$

- The empirical distribution of $\{\gamma^{(1)}, \ldots, \gamma^{(S)}\} \rightarrow p(\gamma\,|\,y^n)$ where $\gamma = g(y)$.