STAT 3006: TUTORIAL4

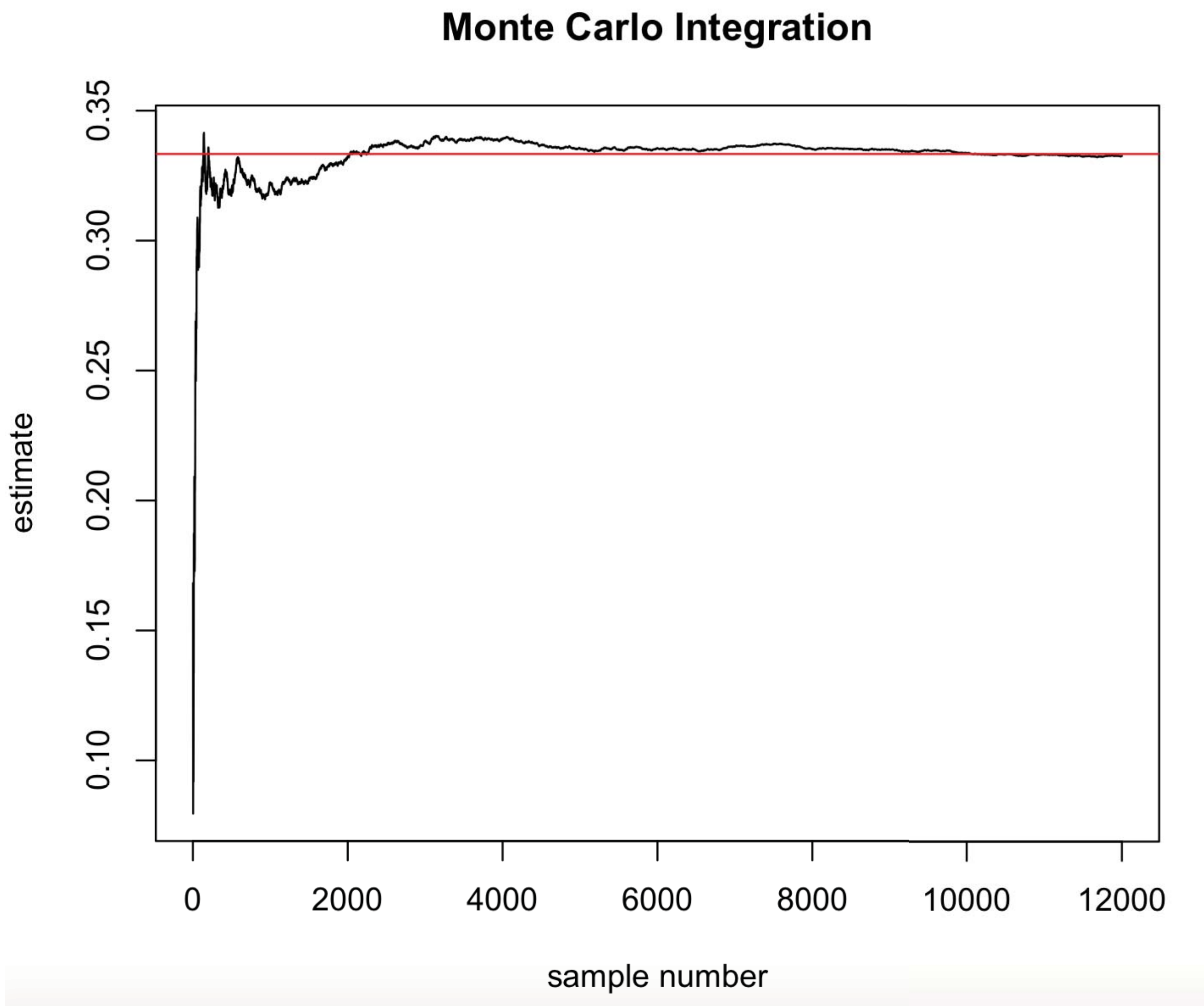
1. Monte Carlo Integration
2. Importance Sampling

MONTE CARLO METHOD

- Generally, Monte Carlo method is using randomness to solve statistical or deterministic problems.
  - Monte Carlo is an area of Monaco.
  - Famous for Monte Carlo casino (opening date: 1863).
  - The concept "Probability" comes from gambling there.
- Application:
  - estimate pi
    - Distribute beans on a square
    - Buffon's needle
  - calculate integral
    - sampling

MONTE CARLO INTEGRATION

- Calculate $\int_0^1 x^2\,dx$ using randomness.
- Law of large number: $\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n}$ $\mathrm{i.i.d.} \sim \mathrm{f(x)}$, $\lim \frac{h(x_1)+\ldots+h(x_n)}{n} = \int h(x)f(x)dx$ almost surely.
- $\int_0^1 x^2\,dx = \int_{-\infty}^{+\infty} x^2 I_{[0,1]}(x)dx$.
- Draw samples $X_1, X_2, \ldots, X_n$ from unif$[0,1]$, Monte Carlo estimation is $\frac{1}{n}\sum X_i^2$.


Monte Carlo Integration
estimate
0.10
0.15
0.20
0.25
0.30
0.35
0
2000
4000
6000
8000
10000
12000
sample number

MONTE CARLO INTEGRATION

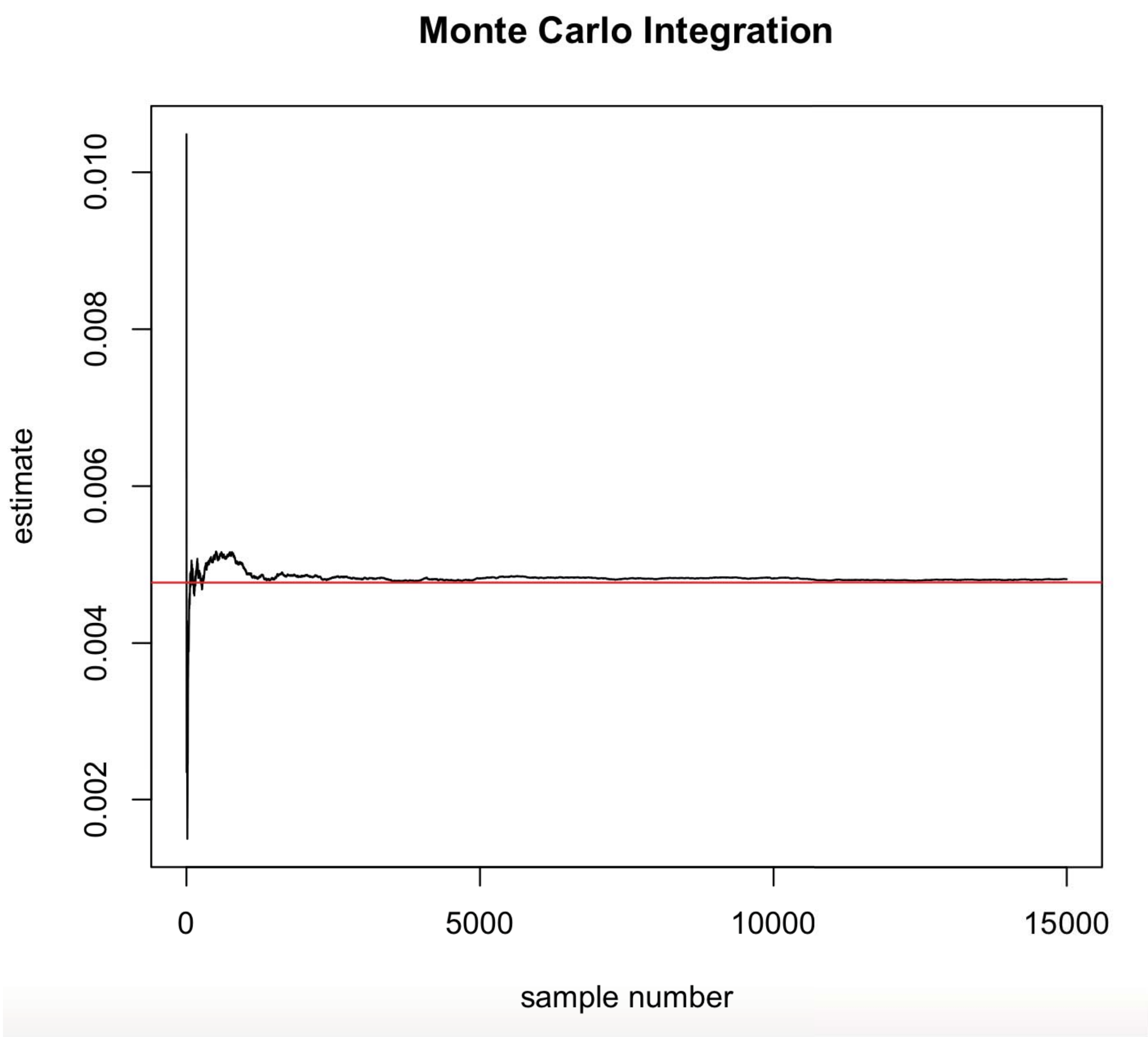
- Example (a)
- How about calculating $\int_{2.5}^{+\infty} \sin(x)\, e^{-\frac{x^2}{2}} dx$?
- $\int_{-\infty}^{+\infty} \sin(x) \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} I_{[2.5,+\infty)}(x)}{1-\boldsymbol{\Phi}(2.5)} dx * \sqrt{2\boldsymbol{\pi}}(1 - \boldsymbol{\Phi}(2.5))$
- Therefore, we have to sample $x_1, \ldots, x_n$ from the truncated normal distribution.
- Monte Carlo estimate is $\frac{\sqrt{2\boldsymbol{\pi}}(1-\boldsymbol{\Phi}(2.5))}{n} \sum \sin(x_i)$
- Accept-reject method.
  - The proposal distribution is Exp(2.5) + 2.5
  - Sometimes, the acceptance rate is not high. For example, draw 1000 samples, only 600 samples are accepted and other 400 samples are useless.

IMPORTANCE SAMPLING

- Notice that $\int h(x)f(x)dx = \int \frac{h(x)f(x)}{g(x)} g(x)dx$, where $g(x)$ is a pdf function.
- Draw i.i.d. samples $\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n}$ from g(x).
- By law of large numbers, $\lim \frac{\frac{h(x_1)f(x_1)}{g(x_1)} + \dots + \frac{h(x_n)f(x_n)}{g(x_n)}}{n} = \int \frac{h(x)f(x)}{g(x)} g(x)dx = \int h(x)f(x)dx.$
- Monte Carlo estimate is $m(x) := \frac{\frac{h(x_1)f(x_1)}{g(x_1)} + \dots + \frac{h(x_n)f(x_n)}{g(x_n)}}{n}$. All samples from g(x) are used.
- However, there is no free lunch. We must put some requirements.
  - Sampling from g(x) is easy. (the same as accept-reject method)
  - Var(m(x)) must be finite.
    - $\frac{f(x)}{g(x)} < M$ for $\forall x$.
    - $E_g h^2 < \infty$
    - g has a larger domain than that of f.

IMPORTANCE SAMPLING

- *Example (a) cont.*
- $\int_{-\infty}^{+\infty} \frac{\mathrm{h(x)f(x)}}{g(x)} g(x)dx * \sqrt{2\boldsymbol{\pi}}(1 - \boldsymbol{\Phi}(2.5))$
- $h(x) = \sin(\mathrm{x}), \mathrm{f(x)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} I_{[2.5,+\infty)}(x)}{1 - \boldsymbol{\Phi}(2.5)}$ (truncated normal), $\mathrm{g(x)} = 2.5e^{-2.5(x-2.5)} I_{[2.5,+\infty)}(x)$ (shifted exponential distribution)
- We can show that $\frac{f(x)}{g(x)} < \frac{e^{-\frac{2.5^2}{2}}}{2.5\sqrt{2\boldsymbol{\pi}}(1-\boldsymbol{\Phi}(2.5))}$, $E_g h^2(x) \leq 1$ satisfying the requirements.
- Sample $\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n}$ from $2.5e^{-2.5(x-2.5)} I_{[2.5,+\infty)}(x)$
- Monte Carlo estimate is $e^{-2.5^2} \frac{\sum \ \ sin(x_i) e^{-\frac{x_i^2}{2} + 2.5x_i}}{2.5n}$.


Monte Carlo Integration
estimate
0.002
0.004
0.006
0.008
0.010
0
5000
10000
15000
sample number