A Guaranteed Automatic Integration Library for Monte Carlo Simulation

Lan Jiang

Department of Applied Mathematics
Illinois Institute of Technology
Email: ljiang14@hawk.iit.edu

Joint work with Prof. Fred J. Hickernell, Dr. Sou-cheng Choi, Yuhan Ding, Lluis Antoni Jimenez
Rugama, Xin Tong, Yizhi Zhang and Xuan Zhou

July 8, 2014
Guaranteed Automatic Integration Library (GAIL)

GAIL is a suite of algorithms for integration problems in one and many dimensions, and whose answers are guaranteed to be correct.

GAIL is created, developed, and maintained by Fred Hickernell (Illinois Institute of Technology), Sou-Cheng Choi (NORC at the University of Chicago and IIT), and their collaborators including Yuhan Ding (IIT), Lan Jiang (IIT), Lluís Antoni Jiménez Rugama (IIT), Xin Tong (IIT), Yizhi Zhang (IIT), and Xuan Zhou (IIT).

To download the latest version of GAIL, follow one of the links below to:

Get zip file OR run the MATLAB installation script

Download zip file

Download MATLAB script
Guaranteed Automatic Integration Library (GAIL):

https://code.google.com/p/gail/ is a suite of algorithms for integration problems in one and many dimensions, whose answers are guaranteed to be correct. It contains four algorithms in its latest release [Choi, Ding, Hickernell, Jiang, Zhang, MATLAB software, 2014]:

- meanMC: Monte Carlo method for evaluating the mean of a random variable.
- cubMC: Monte Carlo method for numerical integration.
- funappx: univariate function recovery.
- integral: trapezoidal rule approximation of one-dimensional integrals.
Guaranteed Automatic Integration Library (GAIL):
https://code.google.com/p/gail/ is a suite of algorithms for integration problems in one and many dimensions, whose answers are guaranteed to be correct. It contains four algorithms in its latest release [Choi, Ding, Hickernell, Jiang, Zhang, MATLAB software, 2014]:

- **meanMC**: Monte Carlo method for evaluating the mean of a random variable.
Guaranteed Automatic Integration Library (GAIL): https://code.google.com/p/gail/ is a suite of algorithms for integration problems in one and many dimensions, whose answers are guaranteed to be correct. It contains four algorithms in its latest release [Choi, Ding, Hickernell, Jiang, Zhang, MATLAB software, 2014]:

- **meanMC**: Monte Carlo method for evaluating the mean of a random variable.
- **cubMC**: Monte Carlo method for numerical integration.

My collaborators would also give talks during SIAM AM14, Fred Hickernell (Wed. 4:00) Yuhan Ding (Wed. 5:30) Llus Antoni Jimnez Rugama (Tue. 5:40) Yizhi Zhang (TR 9:30) Xuan Zhou (Fr. 8:30)
Guaranteed Automatic Integration Library (GAIL):

https://code.google.com/p/gail/ is a suite of algorithms for integration problems in one and many dimensions, whose answers are guaranteed to be correct. It contains four algorithms in its latest release [Choi, Ding, Hickernell, Jiang, Zhang, MATLAB software, 2014]:

- **meanMC**: *Monte Carlo method for evaluating the mean of a random variable.*
- **cubMC**: *Monte Carlo method for numerical integration.*
- **funappx**: *univariate function recovery.*

Future work

My collaborators would also give talks during SIAM AM14, Fred Hickernell (Wed. 4:00) Yuhan Ding (Wed. 5:30) Llus Antoni Jimnez Rugama (Tue. 5:40) Yizhi Zhang (TR 9:30) Xuan Zhou (Fr. 8:30)
Guaranteed Automatic Integration Library (GAIL):

https://code.google.com/p/gail/ is a suite of algorithms for integration problems in one and many dimensions, whose answers are guaranteed to be correct. It contains four algorithms in its latest release [Choi, Ding, Hickernell, Jiang, Zhang, MATLAB software, 2014]:

- meanMC: Monte Carlo method for evaluating the mean of a random variable.
- cubMC: Monte Carlo method for numerical integration.
- funappx: univariate function recovery.
- integral: trapezoidal rule approximation of one-dimensional integrals.

Future work

My collaborators would also give talks during SIAM AM14, Fred Hickernell (Wed. 4:00) Yuhan Ding (Wed. 5:30) Llus Antoni Jimnez Rugama (Tue. 5:40) Yizhi Zhang (TR 9:30) Xuan Zhou (Fr. 8:30)
Guaranteed Automatic Integration Library (GAIL):

https://code.google.com/p/gail/ is a suite of algorithms for integration problems in one and many dimensions, whose answers are guaranteed to be correct. It contains four algorithms in its latest release [Choi, Ding, Hickernell, Jiang, Zhang, MATLAB software, 2014]:

- **meanMC**: Monte Carlo method for evaluating the mean of a random variable.
- **cubMC**: Monte Carlo method for numerical integration.
- **funappx**: univariate function recovery.
- **integral**: trapezoidal rule approximation of one-dimensional integrals.

The next release is scheduled in Sep. 2014, more algorithms will be added.

My collaborators would also give talks during SIAM AM14, Fred Hickernell (Wed. 4:00) Yuhan Ding (Wed. 5:30) Llus Antoni Jimnez Rugama (Tue. 5:40) Yizhi Zhang (TR 9:30) Xuan Zhou (Fr. 8:30)
What is in GAIL

**Guaranteed Automatic Integration Library (GAIL)**

- **meanMCBernoulli**
- **Numerical Experiments**
- **Future work**

**What is in GAIL**

- **Algorithms**
  - meanMC_g
  - cubMC_g
  - funappx_g
  - integral_g
  - meanMCBernoulli_g
  - funmin_g
  - cubSobol_g
  - cubLattice_g
  - and more

- **Unittests**

- **Documentations**

- **Workouts**

- **OutputFiles**

- **Papers**

- **Utilities**

- **Third party**

**Contact**

ljiang14@hawk.iit.edu

**SIAM AM 2014**
Algorithms to get the reliable error estimation of Bernoulli random variables

How to estimate $\hat{p}$ in order to satisfy the following inequalities?

$$\Pr(|\hat{p} - p| \leq \varepsilon_a) \geq 99\%$$

Absolute error criterion

$$\Pr\left(\frac{|\hat{p} - p|}{p} \leq \varepsilon_r\right) \geq 99\%$$

Relative error criterion
Why not using Central Limit Theorem?

- **Central Limit Theorem** is an asymptotic result and carries no guarantee.
  \[ \Pr(|p - \hat{p}_n| \leq \epsilon) \to 1 - \alpha \text{ as } n \to \infty \Rightarrow N \approx \frac{\Phi^{-1}(1-\alpha)}{4\epsilon^2} \text{ By CLT} \]

- Our method, using Hoeffding’s Inequality, carries probabilistic guarantees with finite sample sizes.
  \[ \Pr(|p - \hat{p}_n| \leq \epsilon) \geq 1 - \alpha \Rightarrow N \geq \frac{\log(2/\alpha)}{2\epsilon^2} \text{ By Hoeffding’s Inequality} \]

The ratio is:
\[ r = \frac{N_{\text{Hoeff}}}{N_{\text{CLT}}} \approx \frac{2 \log(2/\alpha)}{\Phi^{-1}(1-\alpha)} \]

Take
\[ N_{\text{CLT}} = \left\lceil \frac{\Phi^{-1}(1-\alpha)}{4\epsilon^2} \right\rceil \]
\[ N_{\text{Hoeff}} = \left\lceil \frac{\log(2/\alpha)}{2\epsilon^2} \right\rceil \]
**Absolute error criterion:** \( \Pr(|\hat{p} - p| \leq \varepsilon_a) \geq 1 - \alpha \)

User specifies:
- the absolute error tolerance \( \varepsilon_a \geq 0 \)
- the uncertainty \( \alpha \in (0, 1) \)

Using \( n = \left\lceil \frac{\log(2/\alpha)}{2\varepsilon_a^2} \right\rceil \) IID Bernoulli random samples to compute sample mean:
\[
\hat{p} = \frac{1}{n} \sum_{i=1}^{n} Y_i.
\]

**Theorem**

Let \( Y \) be a Bernoulli random variable with mean \( p \), it follows that \( \text{meanMCBernoulli.g} \) using sample size \( n \) yields an estimate \( \hat{p} \) which satisfies the fixed width confidence interval condition:

\[
\Pr(|\hat{p} - p| \leq \varepsilon_a) \geq 1 - \alpha.
\]

**Proof.**

Follow directly by Hoeffding’s Inequality.
Relative error criterion: $\Pr \left( \frac{\hat{p} - p}{p} \leq \varepsilon_r \right) \geq 1 - \alpha$

User specifies:
- The relative error tolerance, $\varepsilon_r > 0$,
- The uncertainty, $\alpha \in (0, 1)$,

1. Set $i = 1$ and compute the sample average $\hat{p}_{n_i}$ using $n_i = \left\lceil -\frac{4^i \log(\alpha_i)}{2\varepsilon_r^2} \right\rceil$ IID Bernoulli random samples that are independent of those used to compute $\hat{p}_{n_1}, \ldots, \hat{p}_{n_{i-1}}$, where $\alpha_i = 1 - (1 - \alpha/2)^{2^{-i}}$.

2. Compute $\hat{p}_L = (\hat{p}_{n_i} - \varepsilon_r 2^{-i})^+$ a highly probable lower bound on $p$.
   - If $\hat{p}_L \geq 2\varepsilon_r 2^{-i}$ then $\hat{p}_L$ is large enough, go to step 3.
   - Else, $i = i + 1$, and go back to step 1.

3. Compute $\hat{p}$ using $n = \left\lceil -\frac{\log(\alpha/4)}{2(\hat{p}_L\varepsilon_r)^2} \right\rceil$ IID Bernoulli random sample that is independent of those used to compute $\hat{p}_{n_1}, \ldots, \hat{p}_{n_i}$. Terminate the algorithm.

If the algorithm terminates, then the relative error criterion is satisfied.
Theorem to guarantee the success of \texttt{meanMCBernoulli}. 

Let $Y$ be a Bernoulli random variable with mean $p$, it follows that \texttt{meanMCBernoulli} using sample size $n$ yields an estimate $\hat{p}$ which satisfies the fixed width confidence interval condition:

$$
\Pr\left( \left| \frac{\hat{p} - p}{p} \right| \leq \varepsilon_r \right) \geq 1 - \alpha.
$$

Proof.

\[
\Pr\left( \left| \frac{\hat{p} - p}{p} \right| \leq \varepsilon_r \right) \geq \Pr\left( |\hat{p} - p| \leq \hat{p}_L \varepsilon_r \; \& \; \hat{p}_L \leq p \right)
\geq \Pr\left( |\hat{p} - p| \leq \hat{p}_L \varepsilon_r \; \& \; \hat{p}_L - \varepsilon_r 2^{-i} \leq p \; \forall i \right)
\geq \Pr\left( |\hat{p} - p| \leq \hat{p}_L \varepsilon_r \right) + \Pr\left( \hat{p}_n_i - \varepsilon_r 2^{-i} \leq p \; \forall i \right) - 1
\]

by $\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1$

\[
\geq (1 - \alpha/2) + \prod_{i=1}^{\infty} \left(1 - \frac{\alpha}{2}\right)^{2^{-i}} - 1 \quad \text{By Hoeffding's Ineq and independance}
\]

\[
= 1 - \alpha
\]
Numerical Experiments

Using 500 random samples with true \( p: \log_{10}(p) \sim U[-3, -1] \), we calculate the estimated \( \hat{p} \) to some specified, random chosen, absolute error tolerance and relative tolerance. The figures below shows the results:

![Graphs showing error/tolerance comparison for different p](image)
Future work

- Guarantee the upper bound on cost of meanMCBernoulli\_g.
- Theoretical research of meanMC\_g with relative error criterion.
- Algorithm implementation and GAIL development.
Acknowledgements

- I would like to express my special appreciation and thanks to my advisor Professor Fred J. Hickernel, and all other collaborators.
- I would like to thank my husband Xuan, my two kids Audrey and Alvin, you support and encourage me unconditionally.
- I would also like to express my thank to SIAM AM 2014 organizers for hosting this wonderful event.

Thanks for your attention!
Relative error criterion

Proof.

As for each step
\[ n_i = \left\lceil -\frac{\log(\alpha_i)}{2(\varepsilon_r 2^{-i})^2} \right\rceil \geq -\frac{\log(\alpha_i)}{2(\varepsilon_r 2^{-i})^2} \Rightarrow 1 - e^{-2n_i(\varepsilon_r 2^{-i})^2} \geq 1 - \alpha_i = (1 - \alpha/2)^{2^{-i}}, \]

apply one side Hoeffding’s Inequality:

\[ \Pr(p \geq \hat{p}_{n_i} - \varepsilon_r 2^{-i}) \geq 1 - e^{-2n_i(\varepsilon_r 2^{-i})^2} \geq (1 - \alpha/2)^{2^{-i}} \]

Similarly, as \( n = \left\lceil -\frac{\log(\alpha/4)}{2(\hat{p}_L \varepsilon_r)^2} \right\rceil \geq -\frac{\log(\alpha/4)}{2(\hat{p}_L \varepsilon_r)^2} \Rightarrow 1 - 2e^{-2n(\hat{p}_L \varepsilon_r)^2} \geq 1 - \alpha/2, \) apply two side Hoeffding’s inequality again,

\[ \Pr(|p - \hat{p}_n| \leq \hat{p}_L \varepsilon_r) \geq 1 - 2e^{-2n(\hat{p}_L \varepsilon_r)^2} \geq 1 - \alpha/2 \]
Tools needed to solve the problem

With the help of Bernoulli and Hoeffding, I found a way to solve this problem and win the prize.

- A Bernoulli random variable $Y$ is one that takes on the values 0 or 1 according to
  $$P(Y = 1) = p \quad P(Y = 0) = 1 - p$$
  Thus, $\mathbb{E}(Y) = p$ and $\text{var}(Y) = p(1 - p)$.

- By Hoeffding’s Inequality, if the random variable $Y$ is bounded in $[0, 1]$, and let $Y_1, \cdots, Y_n$ be IID observations such that $E(Y_i) = p$, then the following inequalities hold:
  $$P(p \geq \hat{p}_n + \varepsilon) \leq e^{-2n\varepsilon^2}$$
  $$P(p \leq \hat{p}_n - \varepsilon) \leq e^{-2n\varepsilon^2}$$
  $$P(|\hat{p}_n - p| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$$

lj@hawk.iit.edu
SIAM AM 2014