Monte Carlo with user-specified error

Mark Huber
Fletcher Jones Foundation Associate Professor of Mathematics and Statistics and George R. Roberts Fellow
Claremont McKenna College
16 Aug, 2016

Supported by NSF grant DMS 1418495

Main results

Main results

Theorem (H. 2013)

Given $B_1, B_2, \ldots$ iid Bern(p) where p is unknown, there exists an estimate $\hat{p}_k = \hat{p}_k(B_1, \ldots, B_T)$ such that the distribution of $p/\hat{p}_k \sim$ Gamma$(k, (k-1)c)$ for user-specified k and c, and $\mathbb{E}[T] = k/p$.

Theorem (H. 2016)

Given $A_1, A_2, \ldots$ iid Pois(μ) where μ is unknown, there exists an estimate $\hat{\mu}_k = \hat{\mu}_k(A_1, \ldots, A_T)$ such that the distribution of $\mu/\hat{\mu}_k \sim$ Gamma$(k, (k-1)c)$ for user-specified k and c, and $\mathbb{E}[T] \in [k/\mu, k/\mu + 1]$.

Main results, part II

Theorem (Feng, H., Ruan 2016)

For both of the previous theorems, setting $c = 1$ gives an unbiased estimator. Setting

$$c = \frac{2\epsilon}{(1-\epsilon^2)[\ln(1+\epsilon) - \ln(1-\epsilon)]}$$

makes

$$\mathbb{P}(|\textit{relative error}| > \epsilon) \leq c_1 \cdot c_2^k$$

where c_2 is as small as possible.

Relative error

$$\frac{\hat{a}}{a} - 1$$

By the numbers

Say

$$p = 20\%$$

Suppose want

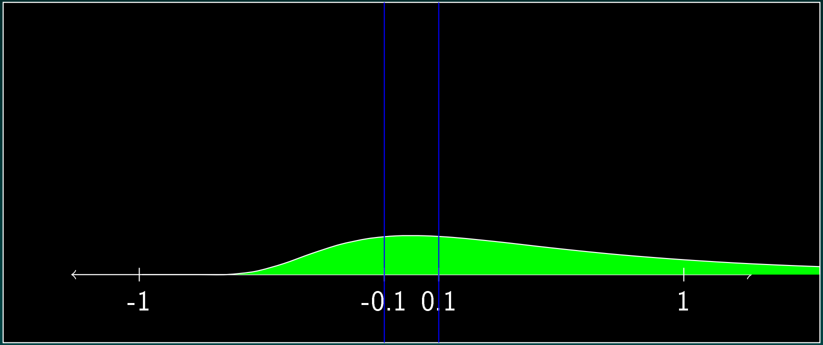
$$|\text{relative error}| \leq 10\%$$

Then need

$$18\% \leq \hat{p} \leq 22\%$$

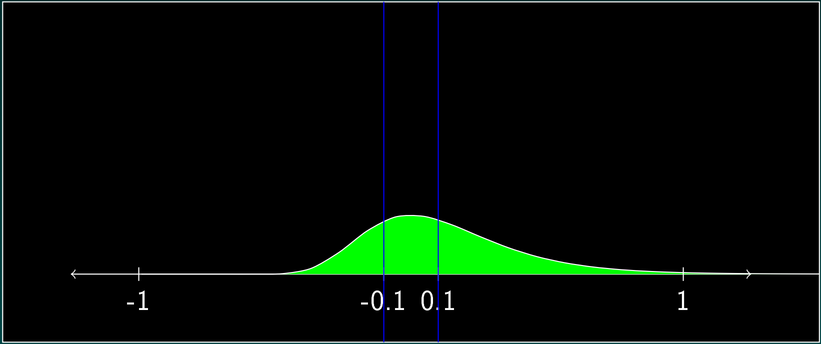
New algorithms know relative error distribution

User set $k = 5$, $c = 1$ $\mathbb{P}(|\text{rel err}| > 0.1) \approx 92.6\%$



New algorithms know relative error distribution

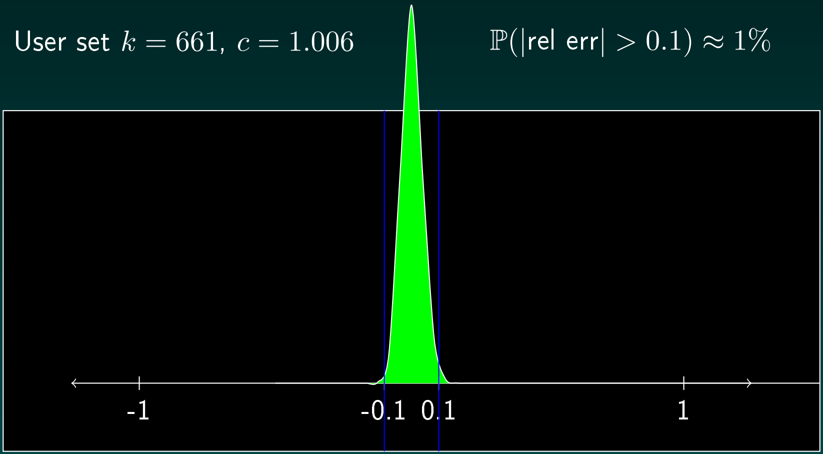
User set $k = 20,\ c = 1$ $\mathbb{P}(|\text{rel err}| > 0.1) \approx 66.1\%$



New algorithms know relative error distribution

User set $k = 661$, $c = 1.006$

$\mathbb{P}(|\text{rel err}| > 0.1) \approx 1\%$



Running time

Lemma

Given k, and $X_1, X_2, \ldots$ iid Bern(p), the expected number of draws used by the algorithm is

$$\frac{k}{\mathbb{E}[X_i]}.$$

Lemma

Given k, and $X_1, X_2, \ldots$ iid Pois(μ), the expected number of draws used by the algorithm is in

$$\left[\frac{k}{\mathbb{E}[X_i]}, \frac{k}{\mathbb{E}[X_i]} + 1\right].$$

A-O-k

Suppose want relative error of 10%

- GBAS (for Bernoulli) and GPAS (for Poisson) user decides k
- Example: if $k = 661$, then

$$\mathbb{P}(\text{relative error} > 0.1) < 0.01$$

- If CLT holds perfectly: data $X_1, X_2, \ldots$ iid $\text{N}(\mu, \mu)$ then

$$k = 664$$

Randomized approximation schemes

Definition

An estimate $\hat{a}$ for a is an (ϵ, δ)-*randomized approximation scheme* if

$$\mathbb{P}\left(\left|\frac{\hat{a}}{a} - 1\right| > \epsilon\right) \leq \delta.$$

Application

Acceptance/rejection integration

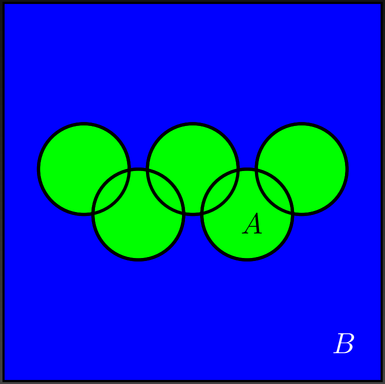
Acceptance/rejection integration used for

Fast approximation of the permanent for very dense problems
M. Huber and J. Law.
Proc. of 19th ACM-SIAM Symp. on Discrete Algorithms, pp 681–689, 2008

Likelihood-based inference for Matérn type-III repulsive point processes
M. L. Huber and R. L. Wolpert
Advances in Applied Probability, Vol 41, No 4, pp. 958–977, 2009

High-Confidence estimator of small s-t reliabilities in directed acyclic networks
R. Zenklusen and M. Laumanns
Networks, Vol 57, No 4, pp. 376–388, 2011

Acceptance/rejection integration



$$X \sim \mathsf{Unif}(B)$$

$$\mathbb{P}(X \in A) = p = \frac{\mathsf{size}(A)}{\mathsf{size}(B)}$$

$$a = \mathsf{size}(A) = \mathsf{size}(B)p$$

$$\hat{a} = \mathsf{size}(B)\hat{p}$$

Multiplication leads to relative error

Multiplication leads to relative error

For estimate

$$\hat{a} = \mathsf{size}(B)\hat{p}$$

it holds that

$$\text{relative error in } \hat{a} = \text{relative error in } \hat{p}$$

How to get relative error small

For basic estimate:

$$\hat{p}_{BE} = \frac{B_1 + \cdots + B_n}{n},$$

- $\mathbb{E}[p_{BE}] = p$, and $\mathrm{SD}(p_{BE}) = \sqrt{p(1-p)/n}$
- To make $\mathrm{SD}(p_{BE}) = \Theta(\epsilon p)$, need $n = \Theta(\epsilon^{-2}/p)$.
- But we don't know p!

DKLR

An optimal algorithm for Monte Carlo estimation
P. Dagum, R. Karp, M. Luby, and S. Ross
Siam. J. Comput., Vol 29, No 5, pp. 1484–1496, 2000

- Idea: Use $\{B_i\}$ to form $\{G_i\}$, where $G_1, G_2, \ldots$ iid Geo(p)
- $\mathbb{E}[G_i] = 1/p$, $\mathrm{SD}(G_i) = (1/p)\sqrt{1-p}$
- Use
$$\hat{p} = \left[\frac{G_1 + \cdots G_k}{k}\right]^{-1}$$
where $k = \Theta(\epsilon^{-2})$ to get relative error below ϵ

Two problems with DKLR

- Biased estimate (in general $\mathbb{E}[1/X] \neq 1/\mathbb{E}[X]$)
- Hard to get correct constant and lower order terms for k

Gamma Bernoulli Approximation Scheme (GBAS)

A Bernoulli mean estimate with known relative error distribution
M. Huber
Random Structures & Alg., arXiv:1309.5413, to appear.

- Use $\{G_i\}$ to form $A_1, A_2, \ldots$ iid Exp(p).
- $\mathbb{E}[A_i] = 1/p$, $\text{SD}(A_i) = 1/p$ (lost factor of $\sqrt{1-p}$)

$$\hat{p}_k = \frac{k-1}{A_1 + \cdots + A_k}$$

Nice properties of GBAS

Unbiased

$$\mathbb{E}[\hat{p}_k] = p$$

Relative error

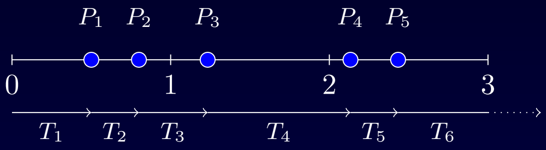
$$p/\hat{p}_k = p/[k-1])A_1 + \cdots (p/[k-1])A_k$$

since $pA_i \sim \mathsf{Exp}(p/(p/(k-1)) \sim \mathsf{Exp}(k-1)$,

$$p/\hat{p}_k \sim \mathsf{Gamma}(k, k-1)$$

How to get exponentials, part 1

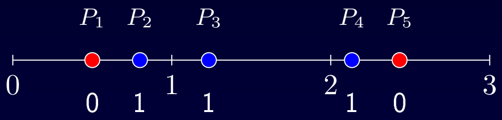
Start with a Poisson point process of rate 1



Here $T_1, T_2, \ldots$ are iid $\text{Exp}(1)$

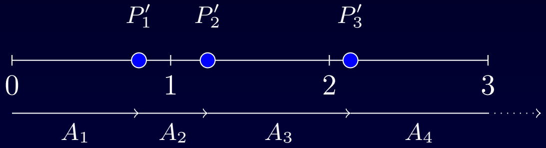
How to get exponentials, part 2

For each point of the process, flip a Bern(p) coin



How to get exponentials, part 3

Only keep those with Bernoulli draw 1



Result is a Poisson point process of rate p

So $A_1, A_2, \ldots$ iid $\mathsf{Exp}(p)$

Geometric sum of exponentials is exponential

Fact

Let $G \sim$ *Geo*(p), *and* $[R|G] \sim$ *Gamma*$(G, 1)$. *Then* $R \sim$ *Exp*(p).

How big should k be?

Bounding tails of gamma distributions for applications
J. Feng, M. Huber, S. Ruan, Y. Zhang
preprint, 2016

Short answer: at most $k = 2\epsilon^{-2}\ln(1/\delta) + 1$.

Long answer

Theorem

For $\epsilon > 0$, $\delta > 0$, *set*

$$c = \frac{2\epsilon}{(1-\epsilon^2)\ln(1+2\epsilon/(1-\epsilon))},$$

and

$$k = 2\epsilon^{-2}\ln(1/\delta) + 1.$$

For $G \sim$ *Gamma*$(k, k-1)$,

$$\mathbb{P}\left(\frac{1}{cG} \in [1-\epsilon, 1+\epsilon]\right) > 1 - \delta\frac{1+\epsilon}{\sqrt{\pi\ln(1/\delta)}}.$$

Application

Mean of a bounded random variable

Converting mean of bounded r.v. to Bernoulli

Lemma

For

$$U \sim \textit{Unif}([0,1]), X \in [0,c]$$

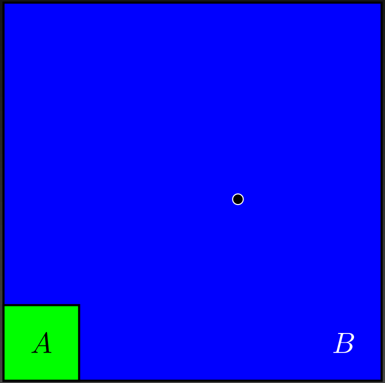
that are independent, then

$$\mathbb{E}[X] = c\mathbb{P}(cU \leq X).$$

Application

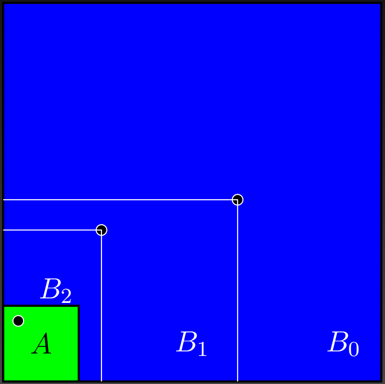
TPA integration

Acceptance/rejection integration



Often the size of A is very small compared to B, making AR slow

TPA



TPA shrinks the set every time we sample. Output X is number of times before point falls into A. In example, $X = 2$.

Nice fact:

$$X \sim \text{Pois}\left(\ln\left(\frac{\text{size}(B)}{\text{size}(A)}\right)\right)$$

Estimating Poisson means

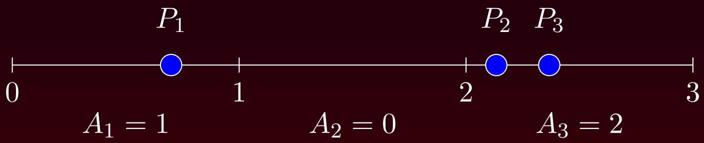
The TPA paper with applications:

Random construction of interpolating sets for high dimensional integration
M. Huber and S. Schott
J. of Applied Probability, arXiv:1112.3692. Vol 51, No 1, pp. 92–105, 2014.

To use TPA to get ratio size(B)/size(A), want estimate for mean of iid Pois(μ) random variables with exact confidence intervals.

How to turn Poissons into Exponentials

- Use $A_1, A_2, \ldots$ iid Pois(μ) for Poisson point process rate μ
- Use $A_i \sim$ Pois(μ) to determine how many points in $[i-1, i]$
- Generate points uniformly in interval



- $P_1, P_2 - P_1, P_3 - P_2, \ldots$ iid Exp(μ)

Running time

- After converting to exponentials, use same as with Bernoullis
- Each Poisson draw gives on average μ exponential draws
- Total Poisson draws for k exponentials is on average in

$$[k/\mu, k/\mu + 1]$$

Pseudocode for GBAS

Gamma_Bernoulli_Approximation_Scheme
Input: k, c *Output:* $\hat{p}_k$
1) $S \leftarrow 0$, $R \leftarrow 0$.
2) Repeat
3) $X \leftarrow \text{Bern}(p)$, $A \leftarrow \text{Exp}(1)$
4) $S \leftarrow S + X$, $R \leftarrow R + A$
5) Until $S = k$
6) $\hat{p}_k \leftarrow (k-1)c/R$

Pseudocode for GPAS

Gamma_Poisson_Approximation_Scheme

Input: k, c *Output:* $\hat{\mu}_k$

$A \leftarrow 0, i \leftarrow 0$

While $A < k$ [Draw k points.]

  $T \leftarrow \text{Pois}(\mu)$

  If $A + T \geq k$ [Then have k points.]

    $T' \leftarrow i + \text{Beta}(k - A, T - (k - A) + 1)$

  $A \leftarrow A + T$, $i \leftarrow i + 1$

$\hat{\mu}_k \leftarrow (k - 1)c/T'$

Main results redux

Theorem

Given $B_1, B_2, \ldots$ *iid Bern*(p) *where* p *is unknown, there exists an estimate* $\hat{p}_k = \hat{p}_k(B_1, \ldots, B_T)$ *such that the distribution of* $p/\hat{p}_k \sim$ *Gamma*$(k, (k-1)c)$ *for user-specified* k *and* c*, and* $\mathbb{E}[T] = k/p$.

Theorem

Given $A_1, A_2, \ldots$ *iid Pois*(μ) *where* μ *is unknown, there exists an estimate* $\hat{\mu}_k = \hat{\mu}_k(A_1, \ldots, A_T)$ *such that the distribution of* $\mu/\hat{\mu}_k \sim$ *Gamma*$(k, (k-1)c)$ *for user-specified* k *and* c*, and* $\mathbb{E}[T] \in [k/\mu, k/\mu + 1]$.

Conclusion

- Given $X_1, X_2, \ldots$ iid either Bernoulli or Poisson with mean a, there is an estimator $\hat{a}$ such that the distribution of the relative error $\hat{a}/a - 1$ is known completely.
- This assists in integration using acceptance/rejection or TPA.
- This also gives exact confidence intervals for the estimate, as well as establishing fast (ϵ, δ)-randomized approximation schemes for many problems.
- By choosing appropriate c, can make the estimate either unbiased, or converge faster than CLT. [This is for Poisson, for Bernoulli, lose a factor of $(1 - p)$.]