Throwing Darts and Needles under Four Configurations: the Uncensored Mean First-Passage-Time of Hitting the k Decimal Digits Value of π

Franc Brglez, Edward Chan, Haiyan Deng, Sanket Goutam, Jun Ma, George Mathew, Yiqi Tang

Franc Brglez	brglez@ncsu.edu	Computer Science, NC State University
Edward Chan	schan3@ncsu.edu	Physics and Computer Science, NC State University
Haiyan Deng	hdeng4@ncsu.edu	Statistics, NC State University
Sanket Goutam	sgoutam@ncsu.edu	Computer Science, NC State University
Jun Ma	jma17@ncsu.edu	Statistics and Bioinformatics, NC State University
George Mathew	george2@ncsu.edu	Computer Science, NC State University
Yiqi Tang	ytang22@ncsu.edu	Statistics, NC State University

ABSTRACT. Throwing needles onto a lined sheet of paper can be considered as the first instance of a stochastic solver that approximates the value of π as reported by Laplace in 1812. This paper examines four distinct stochastic solvers and three stopping criteria to approximate the value of π: `darts`, `needles1`, `needles2`, `needles3`. There is no doubt that increasing the number of throws improves the approximation of π. Also, as a theoretician has already shown for the solver `needles1`, 9.008e+11 is the number of required throws before we can accept with 95% confidence that the correct value of the 5-th decimal place is 9, i.e. the value of π rounded to 5-decimal digits is $\hat{\pi}_5 = 3.14159$. Results in this paper are empirical: in order to approximate $\hat{\pi}_5 = 3.14159$ with the four stochastic solvers, the minimal number of required throws are 275, 355, 452 and the estimated uncensored mean first-passage-time number of throws are 748.8, 3338.9, 8199.2, 9445.2. For the sample size of 1100, the standard errors associated with these mean values are 17.7, 96.2, 241.2, 279.4.

NOTE: The draft of this manuscript has been typeset in LaTeX `tufte-handout` document class (`https://ctan.org/pkg/tufte-latex`). This class is based on Edward R. Tufte. *Beautiful Evidence*. Graphics Press LLC, 2006. (`https://www.edwardtufte.com/tufte/`). The 24-page manuscript features a number of sidenotes and R-generated graphics with empirical asymptotic results for the four stochastic solvers under three stopping criteria. Unfortunately, there are latex issues with posting this version under arxiv and we ask the reader to download the current manuscript (version of December 1, 2018) from this link:

`https://people.engr.ncsu.edu/brglez/publications/OPUS2-2018-pi-tufte-Brglez.pdf`

Currently, only the abstract page is posted under arxiv.org. A complete version of manuscript which will be compatible with latex guidelines under arxiv.org will be posted under `https://arxiv.org/` when it becomes available.

I. Introduction

In 1777, Buffon formulated and solved the problem of finding the probability that a needle of length L thrown onto a horizontal plane ruled with parallel straight lines spaced by a distance $d > L$ will intersect one of these lines. In 1812, Laplace saw this problem in a new light which resulted in a new method of evaluating π. He measured the probability of intersection by throwing the needle onto the ruled paper a very large number of times, recording the fraction of throws resulting in an intersection of the needle with a line. In other words, Laplace applied the frequency estimate of probability. Today, the use of a needle and a lined sheet of paper can be considered as the first instance of a stochastic solver that approximates the value of π as reported by Laplace in 1812.

Initially, introducing Laplace's experiments as the motivational lecture of the project-oriented course on stochastic optimization appeared as the right choice. However, it soon became apparent that additional questions and insights gained during the follow-up lecture could well provide a formal framework for rigorous performance testing of all current and any future stochastic optimization solvers. The goal of this paper is to provide a foundation for such framework.

I.1 About π and the number of decimal digits

The fascination with digits of π has a long history. Examples of π values are listed in Figure 1; details are being explained next.

```
(a) 3.14159 26535 89793 23846 26433
(b) 3.14159 26535 89793 23846 26433 83279 ...
(c) 3.14159 26535 89793 11599 8
(d) 3.14159 26535 89793
```

Figure 1: The digits reported for the value of π. In case (c), digits beyond the 15-th decimal digit are in error. See the adjacent text for more details.

(a) a 20-decimal digit value as reported in Wikipedia;[1]

(b) 100,000 digits of π – as reported by a π aficionado;[2]

(c) a 22-decimal digit value of π which is in error after the 15-th decimal digit. We can replicate this error with the commands

```
> options(digits=22) ; pi
[1] 3.141592653589793115998
```

under the R-shell.[3] We take this limitation into account when programming in R by relying on the 15-decimal digit value of π.

(d) The 15-decimal digit value, $\pi_{15} = 3.141592653589793$, is also used by NASA for the highest accuracy calculations during interplanetary navigation. Paraphrasing the NASA report: *with respect to Voyager 1, about* 12.5 *billion miles away from Earth, the error in calculating the distance to the Voyageur is about* 1.5 *inches.*[4]

[1] Wikipedia, https://en.wikipedia.org/wiki/Pi

[2] 100,000 digits of π , http://www.geom.uiuc.edu/~huberty/math5337/groupe/digits.html

[3] The R Project for Statistical Computing, https://www.r-project.org/

[4] From jpl.nasa.gov/edu/, March 16, 2016, https://www.jpl.nasa.gov/edu/news/2016/3/16/how-many-decimals-of-pi-do-we-really-need/

I.2 Publication Highlights

The scope and the number of citations about π is overwhelming. The informative and popular book by Petr Beckmann is still in print[5], albeit not without errors in history and mathematics[6]. On the

[5] Petr Beckmann. *A History of Pi, Second Edition*. Golem Press, Boulder, Colorado, 1971

[6] Henry W. Gould, AMS Reviews and Descriptions of Tables and Books, Math. Comp. 28 (1974), 325-339, http://www.ams.org/journals/mcom/1974-28-125/S0025-5718-74-99692-6/S0025-5718-74-99692-6.pdf

other end of a spectrum, a formal mathematical as well as historical perspective on π is available in a recent monograph[7]. An amazing closed-form expression, now known as the BPP formula, was discovered only recently[8].

This formula, reproduced in Eq. 2 and analyzed by way of a *first-passage-time* experiment in Figure 2 provides the framework for the construction of this paper. Whereas the experiments in Figure 2 establish in linear time the minimum number of terms to be included in the formula in order to compute π with exactly k decimal digits, we apply the same principle to experimentally determine the mean first-passage-time for four stochastic solvers. Instead of counting the required number of *terms* in Eq. 2, we count the required number of *throws* (of darts or needles) onto a specific geometric configuration.

Articles that are also a source of relevant concepts and notation used in the paper include this list[9].

In 1901, Lazzarini performed a series of needle-throwing experiments and reported an astonishingly accurate value for π[10]. The experiment is now considered a hoax[11].

Conceptual errors in evaluation of geometric probabilities related to π have been discovered and corrected[12].

[7] Jonathan M. Borwein. *The Life of Pi: From Archimedes to ENIAC and Beyond*, chapter From Alexandria, Through Baghdad, pages 531–561. Springer, Berlin, Heidelberg, 2014

[8] David Bailey, Peter Borwein, and Simon Plouffe. On the rapid computation of various polylogarithmic constants. *Mathematics of Computation of the American Mathematical Society*, 66(218):903–913, 1997

[9] N. T. Gridgeman. Geometric Probability and the Number Pi. *Scripta Mathematica*, 25:183–195, November 1960; Michael D. Perlman and Michael J. Wichura. Sharpening Buffon's Needle. *The American Statistician*, 29:157–163, November 1975; Folkmar Bornemann, Dirk Laurie, Stan Wagon, and Jorg Waldvogel. *The SIAM 100-digit challenge: a study in high-accuracy numerical computing*, volume 86. SIAM, 2004; and Enis Sinikrasan. Throwing Buffon's Needle with Mathematica. *The Mathematica Journal*, 11, 2008

[10] M. Lazzarini. Un' applicazione del calcolo della probabilità alla ricerca sperimentale di un valore approsimato di Pi. *Periodico di Matematica*, 4:140–143, 1901

[11] N. T. Gridgeman. Geometric Probability and the Number Pi. *Scripta Mathematica*, 25:183–195, November 1960; and Lee Badger. Lazzarini's Lucky Approximation of Pi. *Mathematics Magazine*, 67(2):83–91, 1994

[12] Lee L Schroeder. Buffon's needle problem: An exciting application of many mathematical concepts. *Mathematics Teacher*, 67(2):183–186, 1974; and Barry J Arnow. On Laplace's extension of the Buffon needle problem. *The College Mathematics Journal*, 25(1):40–43, 1994

ADDITIONAL SECTIONS ARE ORGANIZED AS FOLLOWS:

II. Background and Motivation:
two formulas about π, formalizing the notation, definitions of four stochastic solvers that converge, with increasing the number of throws, to rounded values of π_k, with k denoting the number of decimal digits.

III. Stopping the Throws and Counting the Misses:
R-code templates that implement three criteria to stop the throws, experiments with four stochastic solvers that generate empirical cumulative distribution functions (ECDFs) and asymptotic models for each solver under three stopping criteria.

IV. Aggregates of Mean First-Passage-Times with Four Solvers

V. Summary and Conclusions

VI. Appendix:
On Limitations of the Stopping Criterion `plain`

II. Background and Motivation

WE ORGANIZE THIS SECTION into five subsections:
(1) formulas that converge to π;
(2) notation and the asymptotic first-passage experiment;
(3) on four stochastic solvers for π;
(4) on asymptotic variances of stochastic solvers;
(5) on minimal number of required throws.

II.1 Two Formulas as Solvers

In 1666, Newton, used a geometric construction to derive the formula for π[13]:

$$\pi = \frac{3\sqrt{3}}{4} + 24\left(\frac{1}{12} - \frac{1}{5\cdot 2^5} - \frac{1}{28\cdot 2^7} - \frac{1}{72\cdot 2^9} - \frac{5}{704\cdot 2^{11}} - \cdots\right) \tag{1}$$

The evaluation of the first 6 terms of Eq. 1 is shown in Table 1. We find by inspection that only the fourth term rounds to correct 3 decimal digits: $\hat{\pi}_3 = 3.142$.

Formulas such as Eq. 1 have been evolving over centuries. Some of the most significant improvements are relatively recent such as the amazing BPP formula[14]:

$$\pi = \sum_{n=0}^{\infty}\left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6}\right)(1/16)^n \tag{2}$$

The first three values of $\hat{\pi}$ expressed by Eq. 2 are listed in Table 2. For the third value, $\hat{\pi} = 3.1415874$, we find by inspection that this value rounds correctly to values of π with 3, 4, or 5 decimal digits: 3.142, 3.1416, or 3.14159.

[13] Look for Equation (19) under http://mathworld.wolfram.com/PiFormulas.html

numTerms	piHat	piHat - pi15
1	1.29904	-1.84255
2	3.29904	0.15745
3	3.14873	0.00714
4	3.14204	0.00044
5	3.14138	-0.00021
6	3.14130	-0.00029

Table 1: Convergence to π_{15} as we increase the number of terms in Eq. 1.

[14] David Bailey, Peter Borwein, and Simon Plouffe. On the rapid computation of various polylogarithmic constants. *Mathematics of Computation of the American Mathematical Society*, 66(218):903–913, 1997

numTerms	piHat	piHat - pi15
1	3.1333333	-0.008259
2	3.1414225	-0.000170
3	3.1415874	-5.26e-06

Table 2: Convergence to π_{15} as we increase the number of terms in Eq. 2.

II.1 Notation and the Asymptotic First-Passage-Time Experiment

The notation in Table 3 and Eq. 9 supports the comprehensive *asymptotic performance experiment* summarized in Figure 2 as well as all other experiments in this paper. This experiment not only establishes the asymptotic convergence rate of Eq. 2; it also serves as a template for the follow-up experiments with four stochastic solvers. In Eq. 2, we count *the number terms* that are required before we can observe a solution that converges to $\hat{\pi}_k = \texttt{round}(\hat{\pi}, k)$ where k is the number of decimal digits. With any of our stochastic solvers, we count *the number of throws* (of darts or needles) that are required before we can observe, *for the first time*, the value of $\hat{\pi}_k$ fall within the range of $\pi_k \pm \Delta_k$.

The *unrounded errors* shown as differences $\hat{\pi} - \pi_{15}$ in Tables 1 and 2 are useful indicators about the rate of convergence towards π_{15}. However, we shall rely on definitions in Table 3 and definitions grouped with Eq. 9 to describe results of the experiment in Figure 2.

$$\begin{aligned}
\pi_{15} &= 3.141592653589793 && (3)\\
\pi_k &= \texttt{round}(\pi_{15}, k) \quad (\textit{the target value}) && (4)\\
\Delta_k &= \pm 0.5/10^k && (5)\\
\hat{\pi} &= \textit{approximation of } \pi \textit{ by a solver} && (6)\\
\hat{\pi}_k &= \texttt{round}(\hat{\pi}, k) && (7)\\
\epsilon_k &= \hat{\pi} - \pi_k && (8)\\
\epsilon_{kk} &= \texttt{round}(\epsilon_k, k) && (9)
\end{aligned}$$

There are two phases of the experiment in Figure 2. In Phase 1, we implement the R-function `fg_pi_BPP_asym` exactly as shown. When

symbol	name	description
t	`throws`	number of throws (darts or needles)
t_U	`throwsU`	a limit on the number of throws before the next restart
t_{Lmt}	`throwsLmt`	a limit on the total number of throws
m	`misses`	number of misses (darts or needles)
h	`hits`	number of hits ($= t - m$)
ρ	`restarts`	number of restarts, incremented after each completion of t_U throws
k	`decDigits`	number of decimal digits used to round the value of $\hat{\pi}$
π_k	`pi_k`	value of π rounded to k decimal digits
Δ_k	`piTol_k`	tolerance value ($\pm 0.5/10^k$) associated with value of π_k
$\hat{\pi}$	`piHat`	approximate value of π returned by a solver
$\hat{\pi}_k$	`piHat_k`	value of $\hat{\pi}$ rounded to k decimal digits
ϵ_k	`piError_k`	error difference = $\hat{\pi} - \pi_k$
ϵ_{kk}	`piError_kk`	rounded value of ϵ_k. Solver stops *uncensored* if $(\epsilon_{kk} == 0)$ before $t < t_{Lmt}$.
	`piNewton`	solver, defined by Eq. 1
	`piBPP`	solver, defined by Eq. 2
	`darts`	solver, throwing darts onto *a circle*
	`needles1`	solver, throwing needles onto *a single-grid configuration*
	`needles2`	solver, throwing needles onto *a double-grid configuration*
	`needles3`	solver, throwing needles onto *a triple-grid configuration*
$\text{AVar}(\hat{\pi})_i$	`AVar(piHat)_i`	asymptotic variance for solver i in Table 4
$t_{95}(\hat{\pi})_{i,k}$	`t95(piHat)_i,k`	required number of throws for solver i in Table 4 to reach target value of π_k with 95% confidence
$t_{Min_{i,k}}$	`tMin_i,k`	minimal number of throws for solver i that are required to obtain a solution with k decimal digits. We say that the solution is *uncensored* if $\epsilon_{kk} = 0$ before $t < t_{Lmt}$
$b_{Min_{i,k}}$	`bMin_i,k`	a binary sequence of length $t_{Min_{i,k}}$ as a representation of the solution: 0's represent `misses`, 1's represent `hits`

Table 3: Symbols, names, and descriptions that represent key elements of notation used in this paper.

invoked, it creates the table shown in the table next to the function: the column names have been added manually. The computational flow is simple: (a) compute and save 18 terms of the BPP formula in Eq. 2, (b) for each value of decimal digit k, compute π_k and Δ_k, then increment terms of the BPP formula to iteratively compute $\hat{\pi}$, $\hat{\pi}_k$, ϵ_k, and ϵ_{kk}. There are *two conditions* that can stop the iterations: $(\epsilon_{kk} == 0)$ or (`terms == termsLmt`). If $(\epsilon_{kk} == 0)$ before reaching (`terms == termsLmt`), we say that the variable denoted as `terms` is *uncensored*. Otherwise, we say that the variable `terms` is *censored*. Clearly, in the context of this paper, variables that are *censored* cannot provide reliable information about convergence properties of any solver that returns $\hat{\pi}$ to approximate the value of π.

THE BOOK ON *First-Passage Processes*[15] explains that

> ... first passage underlies many stochastic processes in which the event, such as a dinner date, a chemical reaction, the firing of a neutron, or the triggering of a stock option relies on a variable reaching a specified value *for the first time* ...

[15] Sidney Redner. *A Guide to First-Passage Processes*. Cambridge University Press, 2001

```
fg_pi_BPP_asym = function(decDigits=0:15)
{
  # precompute the terms of formula piBPP
  piTerms = NULL              ; termsLmt = 18
  for (i in 1:termsLmt) {
    n = i - 1
    piTerms[i] = (4/(8*n + 1) - 2/(8*n + 4) -
                  1/(8*n + 5) - 1/(8*n + 6))*(1/16)^n
  }
  pi15    = 3.141592653589793  ; # best-known-value for k=15

  for (k in decDigits) {
    pi_k     = round(pi15, k) ;# target value
    piTol_k  = 0.5/10^k
    piHat    = 0 ; terms = 0

    while (1) {
      terms      = terms + 1
      piHat      = piHat + piTerms[terms]
      # completed piHat for 'piBPP'; begin quantization
      piHat_k    = round(piHat, k)
      piError_k = piHat - pi_k
      # quantize error, break when piError_kk = 0
      piError_kk = round(piError_k, k)

      if (piError_kk == 0 )       {isCensored = FALSE ; break}
      if (terms      == termsLmt) {isCensored = TRUE  ; break}
    }
    cat(k, terms, piHat_k, piError_k, piTol_k, isCensored, "\n")
  }
} # fg_pi_BPP_asym
```

The R-code on the left is also a template for the *first-passage-time* experiments with R-code that implements stochastic solvers such as shown in Figure 8. We only need to interchange variables `terms` and `throws` while maintaining the same formulation of the stopping criterion $\epsilon_{kk} == 0$.

k	terms	piHat_k	piError_k	piTol_k
--	-----	-------	---------	-------
0	1	3	0.133	0.5
1	1	3.1	0.0333	0.05
2	2	3.14	0.00142	0.005
3	3	3.142	-0.000413	5e-04
4	3	3.1416	-1.26e-05	5e-05
5	3	3.14159	-2.61e-06	5e-06
6	5	3.141593	-3.55e-07	5e-07
7	6	3.1415927	-4.68e-08	5e-08
8	5	3.14159265	-4.54e-09	5e-09
9	7	3.141592654	-4.27e-10	5e-10
10	7	3.1415926536	-2.71e-11	5e-11
11	8	3.14159265359	-1.03e-12	5e-12
12	9	3.14159265359	-2.48e-13	5e-13
13	9	3.1415926535898	-4.80e-14	5e-14
14	10	3.14159265358979	1.33e-15	5e-15
15	11	3.141592653589793	0	5e-16

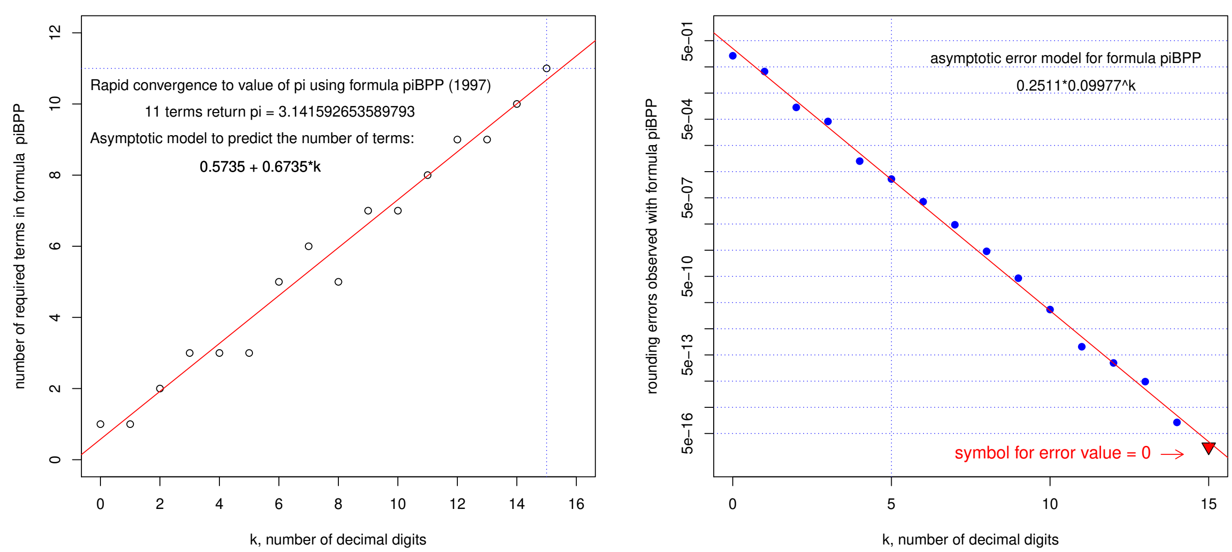


Figure 2: The complete R-code for an asymptotic experiment to verify the efficiency of the BPP formula in Eq. 2. The symbol ▼ denotes $\epsilon_{kk} = 0$ since the error value of 0 cannot be plotted on a log-scale.

As we replace the variable `terms` implied by Eq. 2 with the random variable `throws` implied by *throws* in stochastic solvers, we argue that the iterative search as formulated with Figure 2 can also be viewed as an example of the *first passage process* where the stopping criterion ($\epsilon_{kk} == 0$) is the event that establishes *the first-passage-time* either for terms or throws. In this paper, we link inextricably both the notion of the *uncensored variables* with the notion of *first-passage-time*.

In Phase 2, we extend the R-code in Figure 2 to plot the tabulated values of terms versus decimal digits k and the tabulated values of ϵ_k and Δ_k versus decimal digits k. Notably, only 11 terms are required before reaching the value of π with 15 decimal digits, the most precise representation of π available in R. The asymptotic model in Eq. 10 associates with the BPP formula in Eq. 2:

$$\epsilon_k = 0.2511 * (0.09977)^k \tag{10}$$

This formula is the reference asymptotic model for the experiments with stochastic solvers in this project.

II.3 On Four Stochastic Solvers for π

In contrast to solvers based on formulas that converge to value of π rapidly, the stochastic solvers we consider in this paper converge to π not only at much lower rates but we also observe significant differences in the convergence rate of each solver. The names of the four solvers have already been introduced in Table 3. We now introduce formulas for the estimated value $\hat{\pi}$ of each solver, along with companion illustrations that suggest their configuration and evaluation.

darts

Darts are thrown onto a plane featuring a circle with a unit diameter, see Figure 3. We count the number of throws t and the number of misses m defined by the event of dart not hitting the circle.

We use Eq. 11 to evaluate the estimated value of $\hat{\pi}$.

$$\hat{\pi} = 4 * (1 - m/t) \tag{11}$$

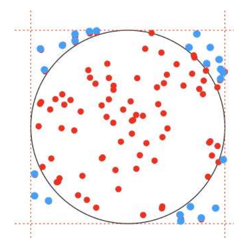


Figure 3: The darts solver.

needles1

Needles of unit length are thrown onto a plane featuring a single grid defined by parallel lines a unit distance apart, see Figure 4. We count the number of throws t and the number of misses m (blue dots) defined by the event of needle not hitting any of the lines.

We use Eq. 12 to evaluate the estimated value of $\hat{\pi}$.

$$\hat{\pi} = 2/(1 - m/t) \tag{12}$$

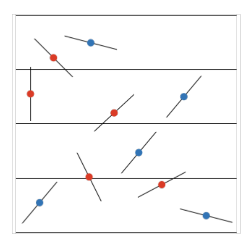


Figure 4: The needles1 solver.

needles2

Needles of unit length are thrown onto a a plane featuring a double grid defined by a grid of unit squares, see Figure 5. We count the number of throws t and the number of misses m defined by the event of needle not hitting any of the lines.

We use Eq. 13 to evaluate the estimated value of $\hat{\pi}$.

$$\hat{\pi} = 3/(1 - m/t) \tag{13}$$

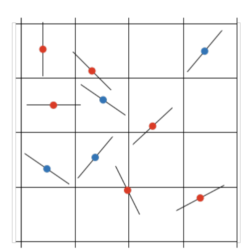


Figure 5: The needles2 solver.

`needles3`

Needles of unit length are thrown onto a a plane featuring a triple grid defined by a grid of equilateral triangles of unit height, see Figure 6. We count the number of throws t and the number of misses m defined by the event of needle not hitting any of the lines.

We use Eq. 14 to evaluate the estimated value of $\hat{\pi}$.

$$\hat{\pi} = 3 * (4 - sqrt(3)/2)/(3 - 2 * m/t) \tag{14}$$

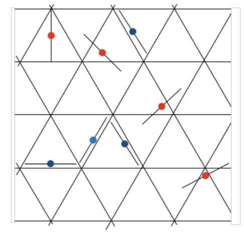


Figure 6: The `needles3` solver.

Each of the four formulas, from Eq. 11 to Eq. 14, plays an important role both in determining *asymptotic variance* of each solver as well as in supporting the search for the *minimal number of throws* that are required before either of the four solvers reaches, *for the first time*, the target value

$$\hat{\pi}_k = \texttt{round}(\hat{\pi}, k)$$

where k denotes the number of decimal digits in $\hat{\pi}_k$.

The next two subsections provide details about the asymptotic variance and the minimal number of throws, respectively.

II.4 *On Asymptotic Variances of Stochastic Solvers*

The asymptotic variances of the four stochastic solvers are summarized in Table 4. The symbol for the asymptotic variance, $\text{AVar}(\hat{\pi})_i$, paraphrases the notation used by Sinikrasan.[16]. Only the variance for solver `darts`, $\text{AVar}(\hat{\pi})_0 = 2.69$, has been derived in this paper. The variances for $\text{AVar}(\hat{\pi})_1 = 5.63$ and $\text{AVar}(\hat{\pi})_2 = 0.4658$ and the basis for the required number of throws, $t_{95}(\hat{\pi})_{i,k}$, originate with Gridgeman.[17] The variance $\text{AVar}(\hat{\pi})_3 = 0.01578$ originates with Perlman and Wichura.[18] The values of $\text{AVar}(\hat{\pi})_1$ and $\text{AVar}(\hat{\pi})_2$ are in agreement with values derived earlier by Gridgeman.

[16] Enis Sinikrasan. Throwing Buffon's Needle with Mathematica. *The Mathematica Journal*, 11, 2008

[17] N. T. Gridgeman. Geometric Probability and the Number Pi. *Scripta Mathematica*, 25:183–195, November 1960

[18] Michael D. Perlman and Michael J. Wichura. Sharpening Buffon's Needle. *The American Statistician*, 29:157–163, November 1975

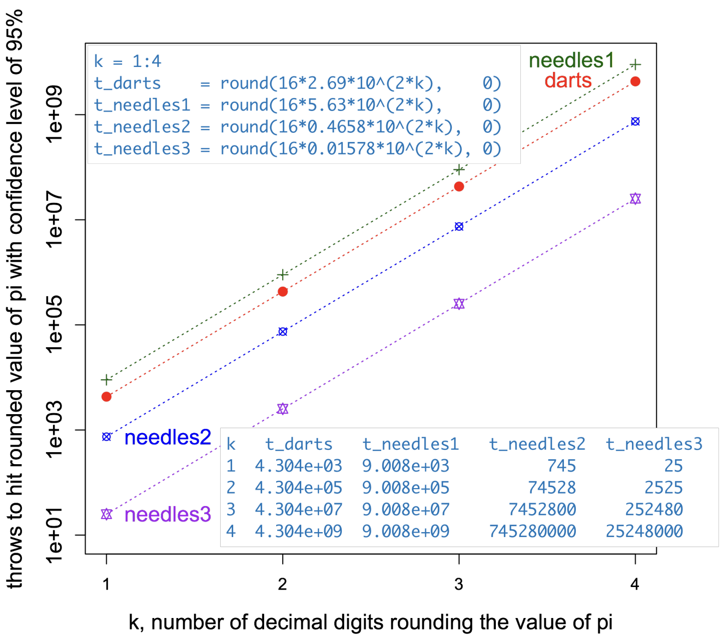
An important question posited by Gridgeman in his 1960 paper is this: "How many casts of the needle must be made before I can confidently accept the k-th decimal place of my estimate of π?" Using current notation, Eq. 15 below paraphrases Gridgeman's Eq. 24 for solver 1:

$$2 * \sqrt{5.63/t_{95}(\hat{\pi})_{1,k}} = 0.5/10^k \tag{15}$$

An asymptotic model, based on solutions of Eq. 15, lists the required number of throws $t_{95}(\hat{\pi})_{i,k}$ in Table 4 and in Figure 7. While we could readily verify this model for $k = 1,\ k = 2$ with solvers `needles2` and `needles3`, the number of required throws $t_{95}(\hat{\pi})_{1,k}$ increases rapidly.

i	solver$_i$	AVar$(\hat{\pi})_i$	$t_{95}(\hat{\pi})_{i,k}$
0	`darts`	$2.69/t$	$\texttt{round}(16 * 2.69 * 10^{(2*k)}, 0)$
1	`needles1`	$5.63/t$	$\texttt{round}(16 * 5.63 * 10^{(2*k)}, 0)$
2	`needles2`	$0.4658/t$	$\texttt{round}(16 * 0.4658 * 10^{(2*k)}, 0)$
3	`needles3`	$0.01578/t$	$\texttt{round}(16 * 0.01578 * 10^{(2*k)}, 0)$

Table 4: Solver-specific symptotic variances (AVar$(\hat{\pi})_i$) and 95% confidence bounds on required throws ($t_{95}(\hat{\pi})_{i,k}$), rounded to k decimal digits.



sampleSize	seed	solver	k	piTol_k	throws	S(%)
100	1066	darts	1	0.05	4304	91
100	1215	darts	1	0.05	4304	96
100	1492	darts	1	0.05	4304	94
100	1066	needles1	1	0.05	9008	95
100	1215	needles1	1	0.05	9008	97
100	1492	needles1	1	0.05	9008	96
100	1066	needles2	1	0.05	745	93
100	1215	needles2	1	0.05	745	97
100	1492	needles2	1	0.05	745	98
100	1066	needles3	1	0.05	25	87
100	1215	needles3	1	0.05	25	92
100	1492	needles3	1	0.05	25	91
100	1066	needles2	2	0.005	74528	95
100	1215	needles2	2	0.005	74528	98
100	1492	needles2	2	0.005	74528	95
100	1066	needles3	2	0.005	2525	95
100	1215	needles3	2	0.005	2525	96
100	1492	needles3	2	0.005	2525	94

Figure 7: A model that predicts the number of throws of dart or needles so that we can accept with 95% confidence that the k-th decimal place of $\hat{\pi}_k$ is the correct value of π.

The adjacent table summarizes results of an experiment with *sampleSize* = 100, repeated for 3 different seeds for the specified number of throws and value of $k = 1$ and 2. The experiment confirms the model predictions well: the column `S(%)` reports the percentage of successes where *success* denotes that the k-th decimal digit has been observed as correct. The fact that for `needles3` and $k = 1$, the observed value of `S(%)` is consistently below 95 can be attributed to Eq. 15: the equation is not being sufficiently precise when throws = 25.

II.5 On Minimal Number of Required Throws

INTUITIVELY, THE NUMBER OF THROWS that are required before a stochastic solver may reach the target value of $\hat{\pi}_k$ depends on the specified value of decimal digits k and the random seed. The larger the value of k, more and more throws are required. A highly conservative estimate is provided by Eq. 15: $O(900, 800)$ throws are required before we can accept with probability of 95% that the 2-nd decimal place is the correct value of π.

However, if we ask, for $k = 2$, "How many throws of the needle (or the dart) must be made before we will compute $\hat{\pi}_k$ for the first time?" we expect the required number of throws to be significantly less than 900,800. See Table 5 for answers: e.g. 11 throws, 4 misses for `needles1` and $k = 2$.

Finding the minimum required number of throws so that $\hat{\pi} = \hat{\pi}_k$ for $k > 6$ is a hard search problem. One formulation is suggested in Table 3: encode a binary sequence of length $L \geq t_{Min_{i,k}}$ with 0's representing `misses`, 1's representing `hits`. For `needles1` and $k = 2$: there are a total $\binom{11}{4} = 330$ solutions, one of them is 01111010110.

In this paper, values in Table 5 are a by-product of experiments discussed in the next section. These include the 're-discovery' of $\hat{\pi} = 355/113$, accurate to 6 decimal places, and known since 500 AD. This approximation is attributed to Zu Chongzhi[19].

[19] Zu Chongzhi, `https://en.wikipedia.org/wiki/Zu_Chongzhi`

		solver = darts piHat = 4*(1 - misses/throws)			solver = needles1 piHat = 2/(1 - misses/throws)		
k	piHat_k	ratios	throws	misses	ratios	throws	misses
1	3.1	28/9	9	2	22/7	11	4
2	3.14	22/7	14	3	22/7	11	4
3	3.142	688/219	219	47	333/106	333	121
4	3.1416	688/219	219	47	688/219	344	125
5	3.14159	355/113	452	97	355/113	355	129
6	3.141593	355/113	452	97	355/113	355	129
..	......						
..	......						
15	3.141592653589793						

		solver = needles2 piHat = 3/(1 - misses/throws)			solver = needles3 piHat = 3*(4 - sqrt(3)/2)/ (3 - 2*misses/throws)	
k	piHat_k	ratios	throws	misses	throws	misses
1	3.1	22/7	22	1	1	0
2	3.14	22/7	22	1	191	1
3	3.142	333/106	111	5	246	1
4	3.1416	355/113	244	11	273	1
5	3.14159	355/113	355	16	275	1
6	3.141593	355/113	355	16	2749	10
..	......					
..	......					
15	3.141592653589793					

Table 5: Denoting k as the number of decimal digits and `misses` as the random variable, this table relates the rounded values of $\hat{\pi}$ to the *minimal* number of throws that are required before either of the four solvers *may* reach, *for the first time*, the target value

$$\hat{\pi}_k = \texttt{round}(\hat{\pi}, k)$$

We use the word *minimal* to denote that the required number throws for *needles3* and $k = 6$ are not necessarily minimum! For problem as formulated in this paper, current computational resources are inadequate to measure *uncensored* performance of *these* stochastic solvers much beyond $k > 6$ within a week of computational effort. However, Figure 2 demonstrates that using Eq. 2 to compute the value of π with high accuracy is both highly efficient and effective. The value of π with 15 decimal digits is reportedly used by NASA for the highest accuracy calculations during interplanetary navigations.

III. *Stopping the Throws and Counting the Misses*

We examine three stopping criteria under the three acronyms: `plain`, `fptC`, and `fptR`. For details about each criterion and the respective R-code templates, see subsection III.1. Subsections that follow summarize a number of experiments with four solvers under each of the three stopping criteria.

III.1 On Stopping Templates

All stochastic solvers are subject to specific stopping criteria. To facilitate correct implementation with each solver, we use three R-code templates in Figure 8, each implementing the solver `darts`:

`plain`:
 number of throws is predetermined,

`fptC`:
 first-passage-time without restarts,

`fptR`:
 first-passage-time with restarts.

Notably, both `fptC` and `fptR` are derived directly from the first-passage-time template that verifies the efficiency of the BPP formula in Figure 2. The formulation of the stopping criterion $\epsilon_{kk} == 0$ has not changed; we only interchanged the variable `terms` with `throws`.

```
## solverName         = darts
## stopping criterion = plain
## (stop if (throws == throwsLmt) )
#
seed      = round(0.5 + 1e9*runif(1))
set.seed(seed) ;# initialize RNG
pi15      = 3.141592653589793
throwsLmt = 1e7 ;# 1e1, 1e2, 1e3, 1e4, ...
throws    = 0 ; misses = 0

while (1) {         ;# count throws
  throws = throws + 1
  x      = runif(1,-1,1)
  y      = runif(1,-1,1)
  if (x^2 + y^2 > 1) {misses = misses + 1}
  piHat     = 4*(1 - misses/throws)
  # completed piHat for 'darts'
  piError   = piHat - pi15
  if (throws == throwsLmt) {break}
}
return( c(piHat, piError, seed) )
```

```
## solverName         = darts
## stopping criterion = fptC
## (first-passage-time, no restarts)
#
seed      = round(0.5 + 1e9*runif(1))
set.seed(seed) ;# initialize RNG
pi15      = 3.141592653589793
k         = decDigits
pi_k      = round(pi15, k)
throwsLmt = 4*round(16*2.69*10^(2*k), 0)
throws = 0 ; misses = 0

while  (1) {        ;# count throws
  throws = throws + 1
  x      = runif(1,-1,1)
  y      = runif(1,-1,1)
  if (x^2 + y^2 > 1) {misses = misses + 1}
  piHat      = 4*(1 - misses/throws)
  # completed piHat for 'darts'; begin quantization
  piHat_k    = round(piHat, k)
  piError_k  = piHat - pi_k
  # quantize and break when piError_kk = 0
  piError_kk = round(piError_k, k)
  if (piError_kk == 0)  {
    isCensored = FALSE ; break
  }
  if (throws  == throwsLmt) {
    isCensored = TRUE  ; break
  }
}
return( c(throws, isCensored,
          piHat_k, piError_k, seed) )
```

```
## solverName         = darts
## stopping criterion = fptR
## (first-passage-time, with restarts)
#
seed       = round(0.5 + 1e9*runif(1))
pi15       = 3.141592653589793
k          = decDigits
pi_k       = round(pi15, k)


throwsLmt  = 4*round(16*2.69*10^(2*k), 0)
throwsU[1] =   9
throwsU[2] =  14
throwsU[3] = 219
throwsU[4] = 219
throwsU[5] = 452
throwsU[6] = 452

throws = 0 ; restarts = -1

while (1) {           ;# count restarts
  restarts = restarts + 1
  set.seed(seed)      ;# initialize RNG
  throws2 = 0  ; misses2 = 0


  while  (1) {        ;# count throws
    throws2 = throws2 + 1
    x       = runif(1,-1,1)
    y       = runif(1,-1,1)
    if (x^2 + y^2 > 1) {misses2 = misses2 + 1}
    piHat      = 4*(1 - misses2/throws2)
    # completed piHat for 'darts'; begin quantization
    piHat_k    = round(piHat, k)
    piError_k  = piHat - pi_k
    # quantize and break when piError_kk = 0
    piError_kk = round(piError_k, k)
    if (piError_kk == 0)  {
      isCensored   =  FALSE ; break
    }
    if (throws2    == throwsU[k]) {
      isCensored   =  TRUE  ; break
    }
  }
  throws = throws + throws2
  if (!isCensored)     {break}
  if (throws  == throwsLmt) {
    isCensored = TRUE ; break
  }
  # new seed for the next restart
  seed = round(0.5 + 1e9*runif(1))
}
return( c(throws, restarts, isCensored,
          piHat_k, piError_k, seed) )
```

Figure 8: Snippets of R code that implement three stopping criteria for the stochastic solver `darts`:
`plain`: *number of throws is predetermined,*
`fptC`: *first-passage-time without restarts,*
`fptR`: *first-passage-time with restarts.*
The same stopping criteria apply to other solvers.

III.2 *Experiments with four solvers stopping under* `plain`

The stopping criterion `plain` is the default stopping criterion for most experiments with stochastic solvers today. Its distinguishing feature is simplicity: one needs only to count the throws (or runtime in seconds) and stop the iterations upon reaching the predetermined number of throws or seconds. However, when used to test performance of two or more solvers, interpretation of observed results may become ambiguous. We demonstrate this ambiguity with a series of experiments indexed by

$$throws \times solvers \times seeds$$

where, expressed in R-code and for $n = sampleSize$:

$$throws = c(10^1, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7)$$

$$solvers = c(\texttt{darts}, \texttt{needles1}, \texttt{needles2}, \texttt{needles3})$$

$$seeds = round(1e9 * runif(n = 100, 0))$$

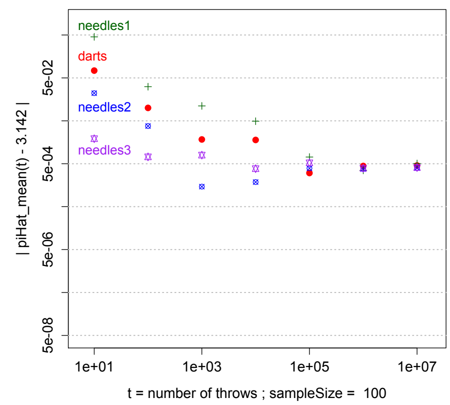
These experiments produce 2800 estimates of $\hat{\pi}$. Several views of these experiments are depicted in Figures 9 and 10. We summarize them next.

At a first glance, the solver performance ambiguities in Figure 9 may appear unexpected. Table 4 lists asymptotic variances of four solvers, with `needles3` exibiting a variance that is two orders of magnitude below `darts`. And yet for $throws = 10^5$, both solvers are returning mean values of $\hat{\pi}$ with nearly the same error, regardless of the reference value of π_k we use: 3.142, 3.14159, or 3.141592653589793 (not shown). Based on the aymptotic variances of four solvers in Table 4, the expected ranking of solvers,

$$\texttt{needles3} > \texttt{needles2} > \texttt{darts} > \texttt{needles1}$$

is being observed for several value of throws – but not for all values. On the other hand, *before* comparing merits of any two stochastic solvers, the question we should ask is this: *"what is the mean number of required throws for each solver to reach the same target range?"* This question is being addressed with the *first-passage-time* stopping criteria `fptC/fptR` in the follow-up sections. For additional arguments that can and do arise when we use `plain` as the stopping criterion, see the Appendix.

Buffon's needles experiments at the top of Figure 10 depict views that are as expected under more traditional analysis. As we increase the number of throws beyond 10^4 we gain better insights by measuring error values rather than values of $\hat{\pi}$ directly. Notably, the plot that depicts sample standard deviation is in a remarkable agreement with predictions of asymptotic variances in Table 4. The histogram classifies 2800 estimates of $\hat{\pi}$ into five categories, defined by the number of decimal digits k. This histogram also reveals a number of useful observations that offer a modicum of explanation for the ambiguities observed in Figure 9: (1) as $k \to 1$, solvers appear more and more equivalent; (2) as $k \to 5$, differentiation between solvers increases dramatically. For errors below 5*e*-06, the tabulated frequency counts in the histogram can also be verified in the adjacent scattergram: $0 + 1 + 1 = 2$ observations for `darts`, $1 + 0 + 0 = 1$



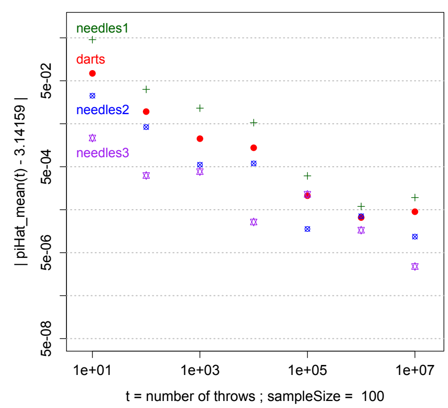
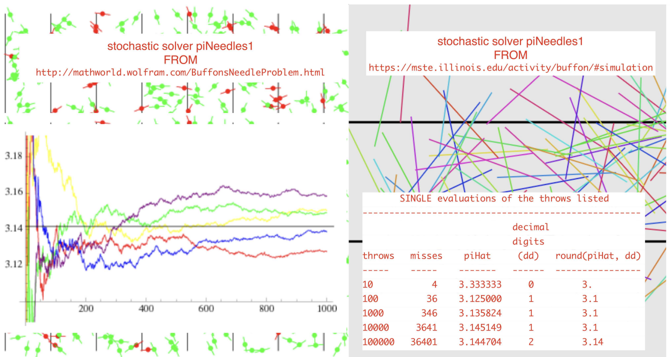


Figure 9: Significant ambiguities *can arise* when we evaluate the performance of four stochastic solvers under a predetermined number of throws (or a predetermined runtime). Later in the paper we demonstrate that under *the first-passage-time stopping criterion*, the solver `needles3` is unambiguously the significantly better solver in comparisons to other solvers. Under the tests shown here, we cannot make such conclusions with high confidence. What reference value should we choose to measure the error of the reported mean value $\hat{\pi}$, how should we chose the 'best' number for throws so we can *reliably* evaluate the performance of four solvers?



These two web sites simulate interactive demonstrations of throwing *Buffon's needles* onto a grid of parallel lines. The convergence towards the value of π is unmistakeable. However, $O(900, 800)$ throws are required before we can accept with high probability that the 2-nd decimal place is the correct value of π. We find the number of throws by solving Eq. 15.

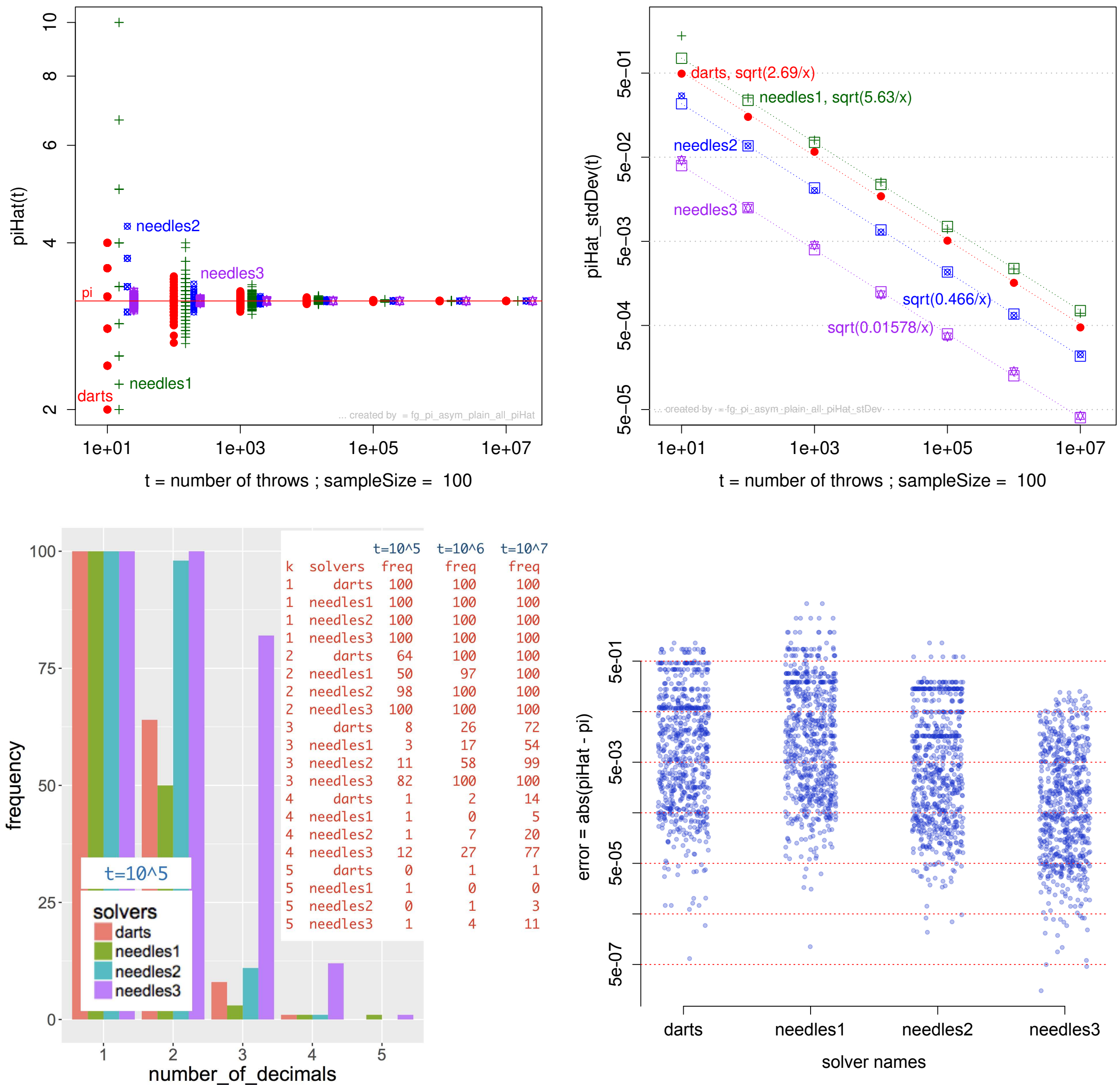


Figure 10: Properties of four stochastic solvers under the stopping criterion `plain`. Each solver stops after exactly $10, 100, 1000, ..., 10,000,000$ throws. A total of $7 * 100 * 4 = 2800$ estimates of $\hat{\pi}$ have been generated.

observations for `needles1`, $0 + 1 + 3 = 4$ observations for `needles2`, $1 + 4 + 11 = 16$ observations for `needles3`. Additional observations of `needles3` are attributed to extensions of experiments with $k < 5$.

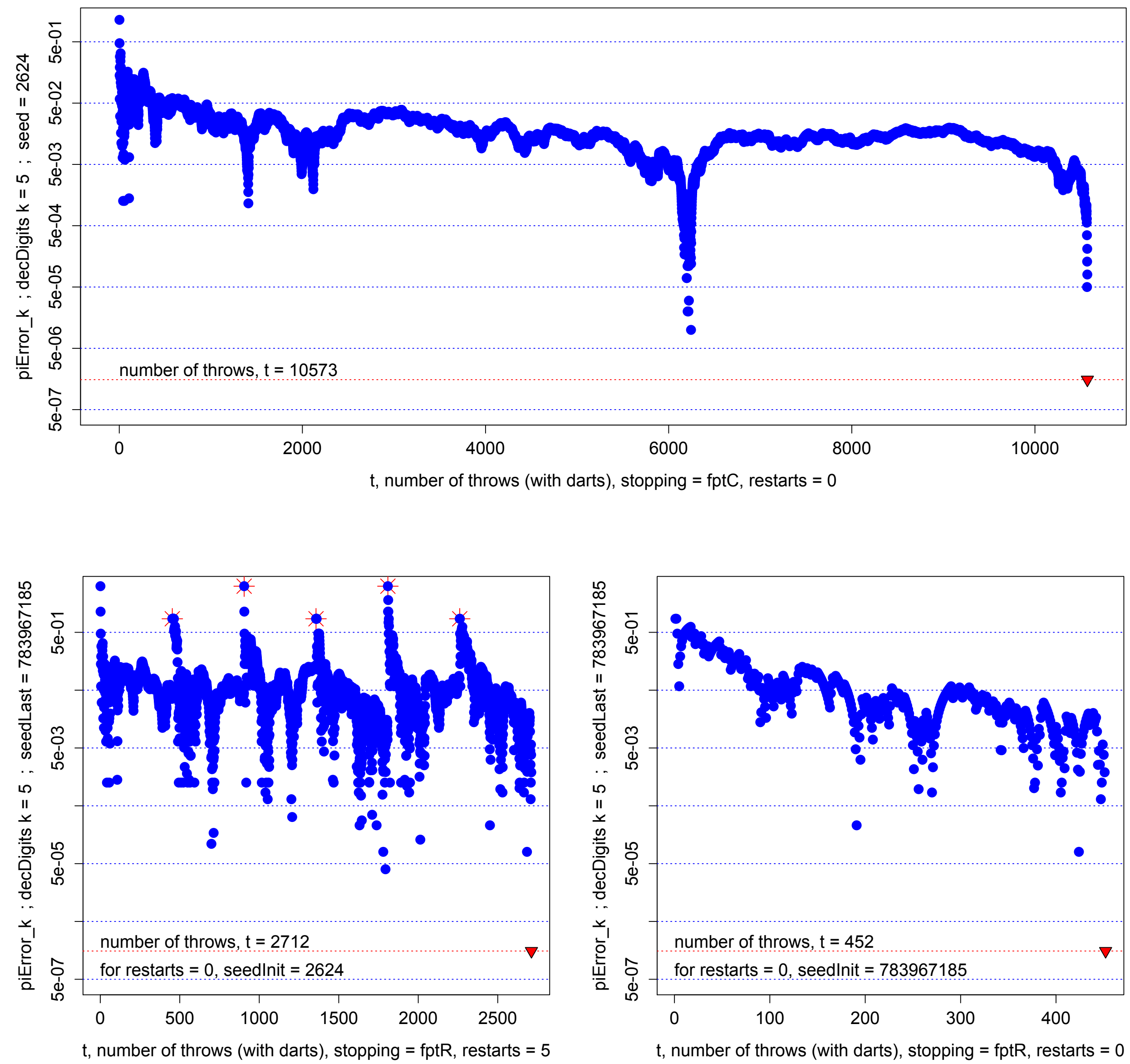


Figure 11: Throwing darts and monitoring the error $\epsilon_k = |\hat{\pi} - \pi_k|$ for $k = 5$ under two stopping criteria, `fptC` (first-passage-time without restarts) and `fptR` (first-passage-time with restarts). Notably, under `fptC` we observe heavy-tail distribution of `throws`. However, under `fptR` we observe near-exponential distribution of `throws`. Under `fptR` we observe the minimal number of throws = 452, where $\epsilon_{kk} = 0$, significantly more frequently than under `fptR`. The symbol ▼ denotes $\epsilon_{kk} = 0$ since the error value of 0 cannot be plotted on a log-scale.

III.3 Experiments with solver `darts` *stopping under* `fptC/fptR`

We divide the experiments with solver `darts` into two part; results are depicted in Figures 11 and 12. The experiment in the top of Figure 11 implements the throws exactly as outlined in the template for stopping criterion `fptC` in Figure 8, given the target value of $\pi_5 = 3.14159$, i.e. for $k = 5$. With the initial seed = 2624, the error $\epsilon_k = \hat{\pi} - \pi_k$ fluctuates until

$$\epsilon_{kk} = \texttt{round}(\epsilon_k, k) = 0$$

i.e. the target value is reached on throw $10,573$. By changing the initial seed to 117691204, the target value is reached only on throw $3,654,750$. For the initial seed = 892288677, the target value is reached in the minimum number of throws: 452. For two initial seeds, 324900668

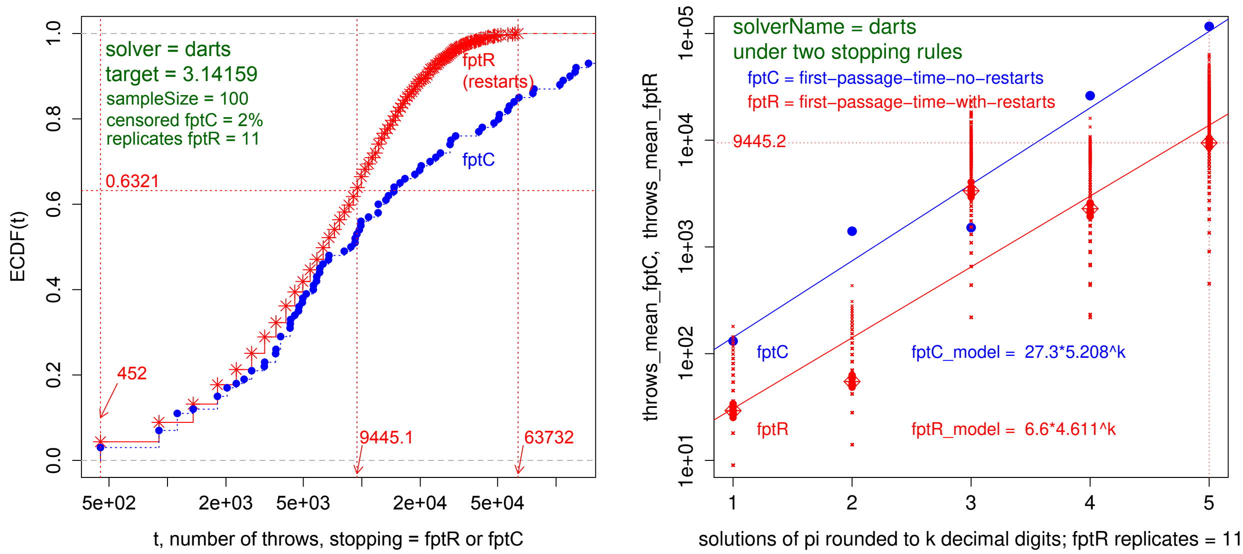


Figure 12: First-passage-time experiments with solver `darts` under two stopping criteria: `fptC` and `fptR`. The results reported for `fptC` are based on a sample size of 100, the results reported for `fptR`, as ECDF and as asymptotic model, are based on a sample size of $11 * 100 = 1100$ where 11 is the number of replicates. The dispersed values of throws are marked with symbol $\times$, the mean value of each replicate are marked with symbol •, the mean value of all 11 replicates is marked with symbol ⊠. The minimal number of throws of 452 in ECDF has been discussed in Table 5.

and 734048780, the target value is not reached since both experiments are being *censored* on reaching the throw limit, $t_{Lmt} = 10^7$. The ECDF for this experiment in Figure 12 suggests a heavy-tail distribution under the stopping criterion `fptC`.

Two experiments in the lower part of Figure 11 implement throws exactly as outlined in the template for `fptR` in Figure 8, With the initial seed = 2624 (the same seed used under `fptC`) , the error $\epsilon_k = \hat{\pi} - \pi_k$ fluctuates between *five random restarts* before reaching the same target value ($\pi_5 = 3.14159$) on throw 2,712. This experiment also reports the last random seed, *seedLast* = 783967185, that invoked the last sequence of throws and reached the target value. By reusing this last seed as the new initial seed in the second experiment, we reach the target value in the minimum number of throws: 452.

Random restarts under the *uncensored first-passage runtime* stopping criterion is one of the heuristics we have used to achieve state-of-the-art solutions for a number of hard combinatorial problems[20]. In each case, we found that the variables we count, such as `throws` in this paper, will have near-geometric or near-exponential distributions.

For a detailed side-by-side comparison of results under two stopping criteria, `fptC` and `fptR`, see Figure 12. The differences in empirical cumulative distribution function (ECDF) under the two stopping criteria increase dramatically with increasing the number of throws; ECDF under `fptR` is clearly near-exponential. The pattern of 5 mean value points for the asymptotic number-of-throws model under `fptR` relates directly to 5 points displayed in Figure 2 for $1 \leq k \leq 5$.

[20] Franc Brglez, Xiao Y. Li, Matthias F. Stallmann, and Burkhard Militzer. Reliable Cost Predictions for Finding Optimal Solutions to LABS Problem: Evolutionary and Alternative Algorithms. In *Fifth Int. Workshop on Frontiers in Evolutionary Algorithms (FEA2003)*, 2003. http://militzer.berkeley.edu/papers/2003-FEA-Brglez-posted.pdf; Borko Bošković, Franc Brglez, and Janez Brest. Low-Autocorrelation Binary Sequences: On Improved Merit Factors and Runtime Predictions to Achieve Them. *Applied Soft Computing Journal – Elsevier*, 2017; and Franc Brglez, Borko Bošković, and Janez Brest. On Asymptotic Complexity of the Optimum Golomb Ruler Problem: From Established Stochastic Methods to Self-Avoiding Walks. *Proceedings of the IEEE Congress on Evolutionary Computation, June 5-8, Donostia – San Sebastian, Spain*, 2017. For a reprint, see https://people.engr.ncsu.edu/brglez/publications.html

III.4 *Three* `needle-type` *solvers stopping under* `fptC/fptR`

Experiments with solver `darts` have provided a method for new experiments and performance comparisons with solvers `needles1`, `needles2`, `needles3`.

```
-----------
darts, fptC (censoredCnt = 2 @10^7)
          median    mean     stDev     SE   cVar  max2mean  min      max
  throws    8457  116606.8  511430.0  51662  4.39     31.3   452  3654750

darts, fptR (censoredCnt = 0)
          median    mean     stDev     SE   cVar  max2mean  min      max
  throws    6328   10174.5   10037.7   1004  0.99     5.86   452    59664
  restarts    13      21.5      22.2    2.2  1.03     6.09     0      131

-------------
needles1, fptC (censoredCnt = 1 @10^7)
          median    mean     stDev     SE   cVar  max2mean  min      max
  throws    9896   63991.2  164914.5  16575  2.58     17.3   355  1105367

needles1, fptR (censoredCnt = 0)
          median    mean     stDev     SE   cVar  max2mean  min      max
  throws    4615    6986.4    6546.2    654  0.94      4.3   355    29820
  restarts    12      18.7      18.4    1.8  0.99      4.4     0       83

-------------
needles2, fptC (censoredCnt = 0)
          median    mean     stDev     SE   cVar  max2mean  min      max
  throws    2707   37552.7  159768.1  15977  4.26     32.4   355  1215393

needles2, fptR (censoredCnt = 0)
          median    mean     stDev     SE   cVar  max2mean  min      max
  throws    2130    3230.5    3281.1    328  1.02      6.7   355    21655
  restarts     5       8.1       9.2    0.9  1.14      7.4     0       60

-------------
needles3, fptC (censoredCnt = 1 @10^7)
          median    mean     stDev     SE   cVar  max2mean  min      max
  throws    1375  124649.4  843521.1  84777  6.77     66.1   275  8240547

needles3, fptR (censoredCnt = 0)
          median    mean     stDev     SE   cVar  max2mean  min      max
  throws     550     693.0     572.4     57  0.83      4.4   275     3025
  restarts     1       1.5       2.1    0.2  1.37      6.6     0       10
```

Table 6: A subset summary of statistics from a total of $5*4*2*100 = 4000$ experiments with five target values of $\hat{\pi}_k$ $(3.1, 3.14, 3.142, 3.1416, 3.14159)$, four solvers (`darts`, `needles1`, `needles2`, `needles3`), under two stopping criteria (`fptC`, `fptR`), and 100 random seeds. The subset is of size 800 for the target value of 3.14159, i.e. for $k = 5$.

sample size for each statistics
$N = 100 -$ `censoredCnt`
standard error
`SE = stDev/sqrt(N-1)`
coefficient of variation
`cVar = stDev/mean`
outlier coefficient
`max2mean = max/mean`

If the distribution of `throws` or `restarts` follows the exponential distribution, the following ratios are invariant:
`median/mean = 0.6931`
`stDev/mean = 1`
`max/mean <= 10` *'almost surely'*
(i.e. with probability of 0.99995)

Each solver considers five target values of $\hat{\pi}_k$, $1 \le k \le 5$:

$$\{3.1,\ 3.14,\ 3.142,\ 3.1416,\ 3.14159\}$$

and stops under `fptC/fptR` criteria.

Statistics in Table 6 summarize such experiments for the case of $k = 5$, i.e. for the target value of 3.14159. For completeness, the table includes the solver `darts`. A quick review of these statistics leads to the following observations about the random variable `throws`:

- stopping under `fptR`, ratios `median/mean`, `stDev/mean`, `max/mean` support the hypothesis that the distribution of `throws` follows the exponential distribution – *regardless of the solver*!
- stopping under `fptC`, distribution of `throws` has a heavier tail than the exponential distribution – *regardless of the solver*!

Results that extend the side-by-side comparisons under two stopping criteria for `darts` in Figure 12 are depicted for `needle-type` solvers in Figure 13. The results reported for `fptC` are based on a sample size of 100, the results reported for `fptR`, in ECDF and in asymptotic

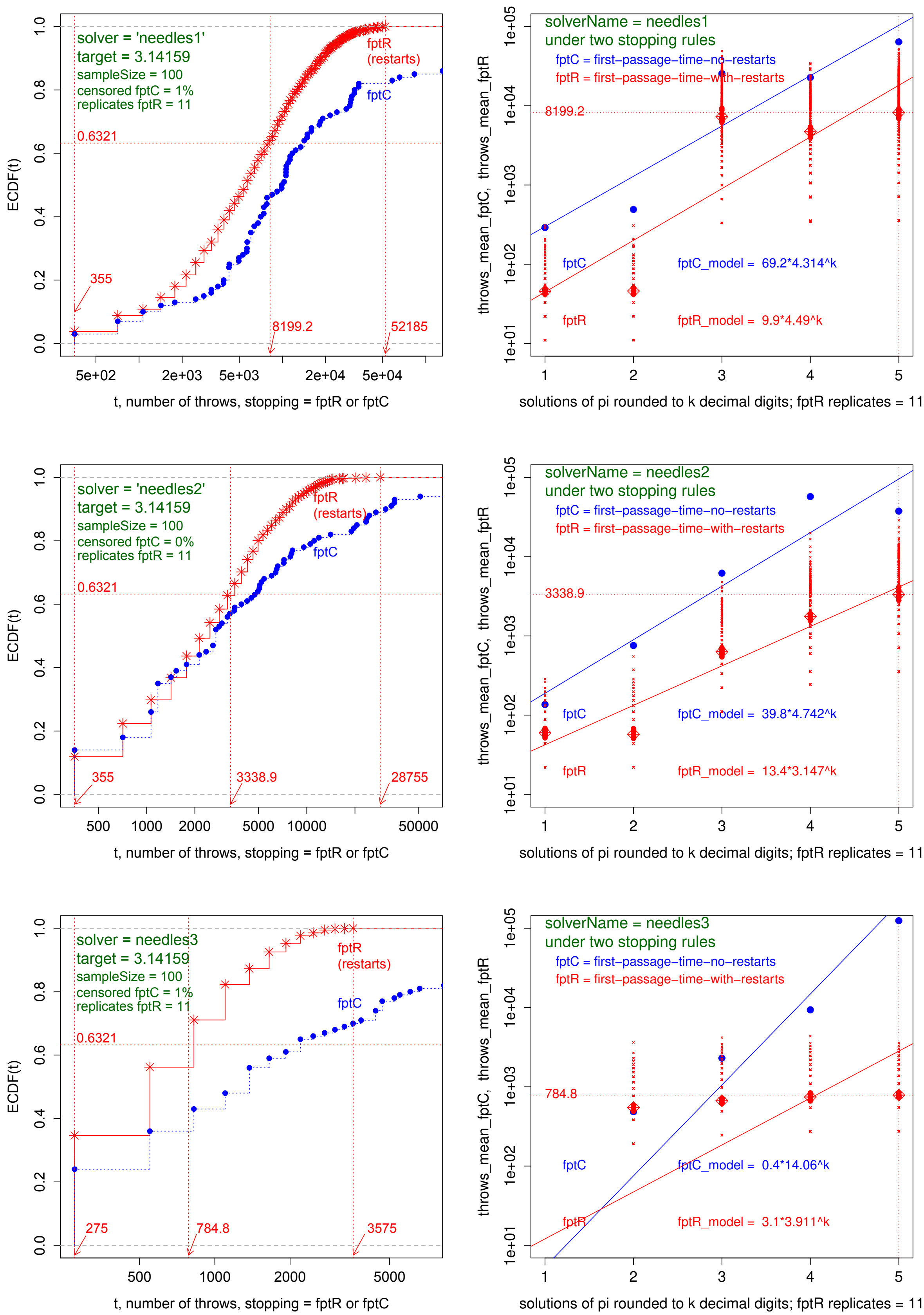


Figure 13: Experiments with three solvers under `fptR` and `fptR`.

model, are based on a sample size of $11 * 100 = 1100$; 11 is the number of replicates. The dispersed values of throws are marked with symbol ×, the mean value of each replicate are marked with symbol •, the mean value of all 11 replicates is marked with symbol ⊠. The minimal number of throws that range from 355 to 275 in the respective ECDFs have been discussed in Table 5.

Additional statistics about experiments that involve 11 replicates and the sample size of $11 * 100 = 1100$ are summarized in Table 7. The main purpose in introducing 11 replicates to experiments under the `fptR` stopping criterion is to convey a measure of reproducibility and precision for each solver. While each solver is returning results that approach the true target value of π, the rates of convergence towards this value differ significantly except for `darts` and `needles1` when sample size ≤ 100. In this context, statistics summarized in Table 6 can only be associated with an instance of a replicate with sample size of 100 under an initial seed value of 1215.

Before proceeding to summarize the result with our four stochastic solvers, we draw attention to the pattern of results with the BPP formula, observed in Figure 2. The pattern of particular interest are five points, for $k = 1, 2, 3, 4, 5$ versus the number of required terms to approximate $\hat{\pi}$ as

$$\{3.1,\ 3.14,\ 3.142,\ 3.1416,\ 3.14159\}$$

Two terms are required to approximate 3.1 or 3.14 respectively. However, the third term alone approximates 3.142, 3.1416, 3.14159. Now, review the mean values of throws associated with $k = 3, 4, 5$ in Figures 12, 13:

k	darts throws	needles1 throws	needles2 throws	needles3 throws
3	3358.1	7292.1	631.3	667.8
4	2283.4	4705.3	1766.1	743.1
5	9445.2	8199.2	3338.9	784.8

We can argue that we are observing a pattern similar to Figure 2: for each solver, the mean value of throws for $k = 5$ is a bound for the expected mean value for $k < 5$. However, the predictive accuracy of the asymptotic model based on $k = 5$ cannot be compared with the empirical model based on $k = 15$ in Figure 2.

```
-----------
darts, fptR (censoredCnt = 0)
            mean     stDev
  throws   8565.4   7142.2   min
  throws   9419.7   9256.7   median
  throws   9445.2   9265.0   mean*
  throws  10454.8  10147.9   max
--------------
needles1, fptR (censoredCnt = 0)
            mean     stDev
  throws   6897.7   7272.2   min
  throws   8271.5   8719.2   median
  throws   8199.2   7998.6   mean*
  throws   9205.2   9279.3   max
--------------
needles2, fptR (censoredCnt = 0)
            mean     stDev
  throws   2815.2   2569.0   min
  throws   3273.1   3060.3   median
  throws   3338.9   3191.9   mean*
  throws   4132.2   4022.5   max
--------------
needles3, fptR (censoredCnt = 0)
            mean     stDev
  throws    739.8    518.1   min
  throws    767.3    584.1   median
  throws    784.8    587.7   mean*
  throws    847.0    550.9   max
```

Table 7: Experiments with four solvers under termination criterion `fptR` and the target value $\hat{\pi}_k = 3.14159$, i.e. with $k = 5$ decimal digits. Each sample is based on 100 random seeds and replicated 11 times. The total sample size for each solver is thus 1100. The rows `min`, `median`, `max` report values of mean and standard deviations from the selection of 11 replicas, each replica with sample size of 100. The rows `mean*` report values of mean and standard deviations computed from 1100 samples contributed by 11 replicas. Plots of ECDFs and asymptotic models for each solver in Figures 12, 13 are based on 1100 samples contributed by all 11 replicas. Similarly, the summary of mean and standard errors in Table 11 are based on 1100 samples for each solver.

IV. *Aggregates of Mean First-Passage-Times with Four Solvers*

ALL THAT REMAINS is to aggregate into a single figure, Figure 14, the most important empirical results reported in several figures and tables earlier: Figures 12, 13, and Tables 6, 7.

The left side of Figure 14 depicts, for each solver, the empirical cumulative distribution functions (ECDFs) for the target value of

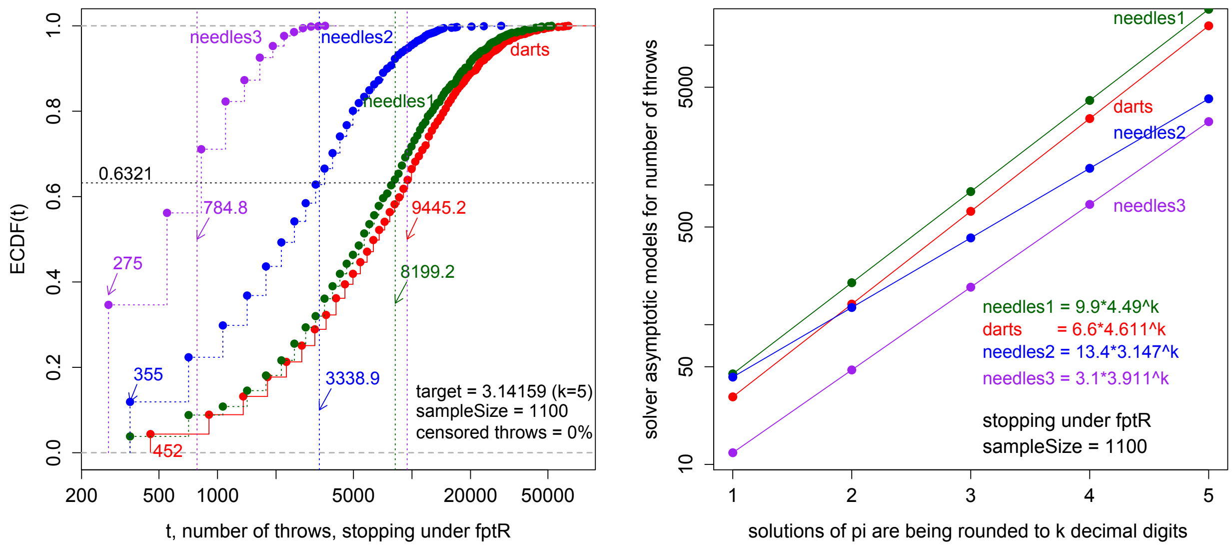


Figure 14: Summary of first-passage-time experiments with four solvers: `darts`, `needles1` , `needles2`, and `needles3`. Each solver returns uncensored number of throws under the same stopping criterion: `fptR` (first-passage-time with restarts). The sample size is $100 * 11 = 1100$ where 11 is the number of replicates.

On the left, we depict empirical cumulative distribution functions (ECDFs) for the target value of 3.14159 ($k = 5$ decimal digits). On the right we depict the asymptotic model that relates the mean number of throws to the number of decimal digits that approximate the value of π.

3.14159 ($k = 5$ decimal digits). The summary of statistics in Table 6 supports the hypothesis that the distribution of `throws` follows the exponential distribution – *regardless of the solver*.

The right side of Figure 14 depicts, for each solver, the asymptotic model that relates the mean number of throws to the number of decimal digits that approximate the value of π. As explained in the closing paragraph of preceding subsection, the predictive accuracy of the asymptotic model based on $k = 5$ cannot be compared with the empirical model based on $k = 15$ in Figure 2. A case in point is the comparison of the mean value of `throws` for solvers `darts` and `needles1`:

from ECDF: mean(darts) > mean(needles1);

from model: mean(darts) < mean(needles1).

All we can say at this point that additional experiments with $k > 5$ are needed to improve the accuracy of the predictive model.

V. Summary and Conclusions

THE INTRODUCTORY SECTION OF THIS PAPER makes an assertion that

> ... insights gained during the follow-up lecture could well provide a formal framework for rigorous performance testing of all current and any future stochastic optimization solvers. The goal of this paper is to provide a foundation for such framework.

The twenty pages that follow, including the appendix, bring together not only the quest for π from two very different directions: from closed-form fomulas to stochastic methods; they also raise new opportunities offered by the very formulation of the stochastic solvers in this paper. The methodology we use to critically examine the role of termination criteria when deciding the trade-offs in their design can be ported to any new stochastic solver, be it in continuous or in discrete domains.

RANDOM RESTARTS UNDER UNCENSORED FIRST-PASSAGE TIME stopping criterion is one of the heuristics we have used to achieve state-of-the-art solutions for a number of hard combinatorial problems[21]. There is every reason to believe that by adapting and adopting these principles for stochastic solvers in a number of different problem domains, further improvements will be observed. In particular, adopting the quantized error as the target value formulated in Figure 8, converging to 0, will further accelerate the solver convergence towards the best-known-values for the problems at hand. Such approaches are under test not only in the continuous domain[22] but also in a number of discrete domains[23].

[21] Borko Bošković, Franc Brglez, and Janez Brest. Low-Autocorrelation Binary Sequences: On Improved Merit Factors and Runtime Predictions to Achieve Them. *Applied Soft Computing Journal – Elsevier*, 2017; and Franc Brglez, Borko Bošković, and Janez Brest. On Asymptotic Complexity of the Optimum Golomb Ruler Problem: From Established Stochastic Methods to Self-Avoiding Walks. *Proceedings of the IEEE Congress on Evolutionary Computation, June 5-8, Donostia – San Sebastian, Spain*, 2017. For a reprint, see `https://people.engr.ncsu.edu/brglez/publications.html`

[22] Franc Brglez. On Uncensored Mean First-Passage-Time Performance Experiments with Multi-Walk in $\mathbb{R}^p$: a New Stochastic Optimization Algorithm. *Invited talk, IEEE Proc. 7th Int. Conf. on Reliability, InfoCom Technologies and Optimization (ICRITO'2018); Aug. 29–31, 2018, Amity University, Noida, India*, 2018. For a preprint, see `https://people.engr.ncsu.edu/brglez/publications.html`

[23] Franc Brglez. On Uncensored Global Stochastic Optimization in $\mathbb{R}^p/\mathbb{D}^p$ and the Efficiency of Multi-Walk Algorithms under Tableau Formulations. *Work in progress. For updates, see `https://people.engr.ncsu.edu/brglez/publications.html`*, 2018

FUTURE WORK

Due to the limited scope of this class project, empirical results in this paper are limited to approximating the asymptotic model of $\hat{\pi}$ to 5 decimal digits only. Unless new stochastic solvers are invented, the current approximations of $\hat{\pi}_5 = 3.14159$ with the four stochastic solvers will stand, for the sample size of 1100, with the estimated uncensored mean first-passage-time number of throws at 748.8, 3338.9, 8199.2, 9445.2 and the respective standard errors at 17.7, 96.2, 241.2, 279.4. Future work may consider a computational project that will extend the asymptotic model of $\hat{\pi}$ to $k > 5$ decimal digits! The R-code and python-code is available on request. For computations with $k > 8$, re-coding in C is recommended.

ACKNOWLEDGEMENTS

The code that created most of the data and illustrations in this article has been written in R and relies on a number of existing R-packages[24].

The layout of the paper, with emphasis on containment of data, illustrations, and sidenotes to the same page, could not have been achieved without the LaTeX package `tufte-handout`[25]. See also the inspiring book on Beautiful Evidence[26].

Kudos to all R-code and latex-code volunteers!

[24] The R Project for Statistical Computing, `https://www.r-project.org/`

[25] tufte-latex – Document classes inspired by the work of Edward Tufte. Download from `https://ctan.org/pkg/tufte-latex?lang=en`

[26] Edward R. Tufte. *Beautiful Evidence*. Graphics Press LLC, 2006. `https://www.edwardtufte.com/tufte/`

VI. *Appendix: on limitations of the stopping criterion* `plain`

Before turning our attention to the title of this appendix, we pause by paying homage to individuals and organizations who,

> circa 1950, began concerted efforts to reach a common agreement or understanding as to the meanings and consequences of using the terms such as "precision" and "accuracy"[27].

Subsequently, the paper by Churchill Eisenhart[28] became the preeminent publication on the subject.

Before getting a step closer to additional context and motivation for this appendix, here are the questions asked by Stanley Rasberry[29]:

> What one question haunts the best of analytical chemists when their day's work is done? Four of the main questions that arise regarding any analytical method are:
> – Is it sensitive enough for the level of detection required?
> – Is it free of interferences for the desired analyte?
> – Is it precise, so that the results are reproducible?
> – Is it accurate, so that the results approach true values?

Such questions are now being asked in undergraduate laboratories[30] where Planck's constant is being measured with relative accuracy and precision. See Table 8 for updates on values of the Planck's constant: 8-significant digits are reported with certainty, with estimated standard deviation of 81 at digits 9 and 10, respectively.

According to web-based resources posted under the Cambridge Energy Landscape Database[31], Lennard-Jones clusters have become a much-studied test system for global optimization methods designed for configurational problems. For an example of a solution, relevant in the context of this paper, see Table 9.

Notably, to find the minimum energy and optimum balance of forces between $N \leq 150$ atoms in a LJ-cluster we need to express not only the optimum energy with accuracy of 9 significant digits but also the xyz solution coordinates with 11 significant digits.

The takeaway from these paragraphs is that both the accuracy and the precision matter when we design and interpret results of lab-based experiments and also computational experiments that rely on 9 to 11 significant digits. Random variables related to solver runtime or counts such as the number of throws will have, in most cases, near-exponential distribution or worse. Sufficient computational resources are required so that statistics used to evaluate solver performances are based strictly on samples that are *not censored*.

We now re-examine the ambiguities that can arise with the stopping criterion `plain` as outlined in Figure 9 earlier. This criterion is the basis for computational experiments whose goal is to rank the performance of stochastic optimization solvers: take S solvers, P problem instances, N random seeds, run each solver under *a fixed*

[27] Prepared by M. Carroll Croarkin, http://nvlpubs.nist.gov/nistpubs/sp958-lide/129-131.pdf

[28] Churchill Eisenhart. Realistic evaluation of the precision and accuracy of instrument calibration systems. *J. of Research of the NBS*, 67:161–187, 1963

[29] Stanley D Rasberry. Accuracy in Analysis: The Role of Standard Reference Materials. *J. of Research of the NBS*, 93(3):213–216, 1988

[30] A. Checchetti and A. Fantini. Experimental Determination of Planck's constant using Light Emitting Diodes (LEDs) and Photoelectric Effect. *World J. of Chemical Ed.*, 3(4):87–92, 2015

```
Planck constant (h)
------------------
Value
     6.626 070 040 x 10-34 J s
Standard uncertainty
     0.000 000 081 x 10-34 J s
Concise form
     6.626 070 040(81) x 10-34 J s
```

Table 8: Plank's constant as posted by NIST in 2014. The standard uncertainty u(y) of a measurement result y is the estimated standard deviation of y. See https://physics.nist.gov/cgi-bin/cuu/Value?h%7Csearch_for=universal_in!

[31] http://doye.chem.ox.ac.uk/jon/structures/LJ/ and the subdirectory on Lennard-Jones Clusters, http://www-wales.ch.cam.ac.uk/CCD.html

```
  N      Energy
 25    -102.372663 (Hoare)
solution
0.6884429539 -0.4902635898 -1.4414708801
1.5396551009 -0.2813139477 -0.5852156965
0.2992192805 -0.8759083275  1.3497840799
 ....
 ....
0.1856536312 -0.1342268229  0.6014654858
0.0479861996  0.5490057018 -0.1816691063
```

Table 9: Global minimum for a Lennard-Jones cluster of $N = 25$ atoms, posted under http://doye.chem.ox.ac.uk/jon/structures/LJ/
Notably: the value of the global optimum is specified with 9 significant digits, the xyz solution coordinates are specified with 11 significant digits!

throws	mean darts	CI95 darts	mean needles3	CI95 needles3	darts - needles3	paired p-value
1e1	308400.000	297503.8 319296.2	314069.026	313307.5 314830.6	-5669.03	0.30558
1e2	312960.000	309979.3 315940.7	314283.994	313996.7 314571.3	-1323.99	0.38240
1e3	313648.000	312588.8 314707.2	314122.606	314038.1 314207.1	-474.61	0.37761
1e4	314161.600	313803.2 314520.0	314155.048	314128.6 314181.5	6.55	0.97122
1e5	314106.760	314011.0 314202.5	314155.082	314146.4 314163.7	-48.32	0.32102

Table 10: This is a summary of a replicated experiment, based on the one in Figure 9. While the stopping criterion `plain` remains unchanged, we consider only the solvers `darts` and `needles3` and monitor the first 9 decimal digits of π only. To make comparisons easier to relate with typical solvers, we rescale the value of π by multiplicative factor of `1e5`. The best-known-value, based on rescaled value of π, rounded to 5 significant digits would thus become 314159.

runtime limit and then, for each solver, tabulate distances from instance BKVs (best-known-values) and a number of related statistics. This is precisely what the experiment in Figure 9 was expected to accomplish for the four stochastic solvers and with the problem instance defined by the *best-known-value* of π.

Ambiguities with experimental results under the stopping criterion `plain` have been observed, on a much larger scale, only recently[32]: Figure 2 of this paper tallies successes with 18 solvers over all 100 runs for each of 48 objective functions. A verbatim quote: “A ‘success’ was defined as a solution less than 0.005 more than the minimum of the objective function between the default bounds." Briefly, solutions where BKV $\leq$ *solution value* $<$ BKV + 0.005 are accepted, solutions not in this range are considered, in our context, as *censored*.

[32] Katharine Mullen. Continuous Global Optimization in R. *Journal of Statistical Software, Articles*, 60(6):1–45, 2014

In other words, the percentage of censored results ranges from `(4800 - 3800)/4800 = 21%` to `(4800 - 1200)/4800 = 75%`. If the error tolerance that defines a ‘success’ is reduced from 0.005 to 0.0005, the percentage of censored results in Figure 2 will get closer to 100%.

Results in Figure 9 are replicated in a new experiment and summarized in Table 10. In comparison with Figure 9, the ambiguities have not changed, they are just better documented. Given these statistics, the mean values of solvers `darts` and `needles3` are indeed equivalent under the null hypothesis – and contrary to our intuitive expectations. We argue that the results reported under the stopping criterion of `plain` are ambiguous because we have not asked the right question *before* we decided to compare performance merits of two stochastic solvers. The right question is this: *“what is the uncensored mean number of required throws for each solver to reach the same target range?”*

solver	mean(throws)	SE(throws)
	sampleSize = 1100	
darts	9445.1	279.4
needles1	8199.2	241.2
needles2	3338.9	96.2
needles3	587.7	17.7

Table 11: The mean number of required throws for each solver to reach the target value of $\texttt{round}(\pi, 5) = 3.14159$. These values are based on 1100 *uncensored samples*, generated under the `fptR` stopping criterion.

Unambiguous answers to this question are summarized in Table 11; all statistics are based on four near-exponential empirical cumulative distribution functions (ECDFs) in Figure 14. The empirical results in this table provide the template for rigorous evaluation of stochastic optimization solvers under test currently[33].

[33] Franc Brglez. On Uncensored Mean First-Passage-Time Performance Experiments with Multi-Walk in $\mathbb{R}^p$: a New Stochastic Optimization Algorithm. *Invited talk, IEEE Proc. 7th Int. Conf. on Reliability, InfoCom Technologies and Optimization (ICRITO’2018); Aug. 29–31, 2018, Amity University, Noida, India*, 2018. For a preprint, see https://people.engr.ncsu.edu/brglez/publications.html

References

[1] Petr Beckmann. *A History of Pi, Second Edition*. Golem Press, Boulder, Colorado, 1971.

[2] Jonathan M. Borwein. *The Life of Pi: From Archimedes to ENIAC and Beyond*, chapter From Alexandria, Through Baghdad, pages 531–561. Springer, Berlin, Heidelberg, 2014.

[3] David Bailey, Peter Borwein, and Simon Plouffe. On the rapid computation of various polylogarithmic constants. *Mathematics of Computation of the American Mathematical Society*, 66(218):903–913, 1997.

[4] N. T. Gridgeman. Geometric Probability and the Number Pi. *Scripta Mathematica*, 25:183–195, November 1960.

[5] Michael D. Perlman and Michael J. Wichura. Sharpening Buffon's Needle. *The American Statistician*, 29:157–163, November 1975.

[6] Folkmar Bornemann, Dirk Laurie, Stan Wagon, and Jorg Waldvogel. *The SIAM 100-digit challenge: a study in high-accuracy numerical computing*, volume 86. SIAM, 2004.

[7] Enis Sinikrasan. Throwing Buffon's Needle with Mathematica. *The Mathematica Journal*, 11, 2008.

[8] M. Lazzarini. Un' applicazione del calcolo della probabilità alla ricerca sperimentale di un valore approsimato di Pi. *Periodico di Matematica*, 4:140–143, 1901.

[9] Lee Badger. Lazzarini's Lucky Approximation of Pi. *Mathematics Magazine*, 67(2):83–91, 1994.

[10] Lee L Schroeder. Buffon's needle problem: An exciting application of many mathematical concepts. *Mathematics Teacher*, 67(2):183–186, 1974.

[11] Barry J Arnow. On Laplace's extension of the Buffon needle problem. *The College Mathematics Journal*, 25(1):40–43, 1994.

[12] Sidney Redner. *A Guide to First-Passage Processes*. Cambridge University Press, 2001.

[13] Franc Brglez, Xiao Y. Li, Matthias F. Stallmann, and Burkhard Militzer. Reliable Cost Predictions for Finding Optimal Solutions to LABS Problem: Evolutionary and Alternative Algorithms. In *Fifth Int. Workshop on Frontiers in Evolutionary Algorithms (FEA2003)*, 2003. http://militzer.berkeley.edu/papers/2003-FEA-Brglez-posted.pdf.

[14] Borko Bošković, Franc Brglez, and Janez Brest. Low-Autocorrelation Binary Sequences: On Improved Merit Factors and Runtime Predictions to Achieve Them. *Applied Soft Computing Journal – Elsevier*, 2017.

[15] Franc Brglez, Borko Bošković, and Janez Brest. On Asymptotic Complexity of the Optimum Golomb Ruler Problem: From Established Stochastic Methods to Self-Avoiding Walks. *Proceedings of the IEEE Congress on Evolutionary Computation, June 5-8, Donostia – San Sebastian, Spain*, 2017. For a reprint, see https://people.engr.ncsu.edu/brglez/publications.html.

[16] Franc Brglez. On Uncensored Mean First-Passage-Time Performance Experiments with Multi-Walk in $\mathbb{R}^p$: a New Stochastic Optimization Algorithm. *Invited talk, IEEE Proc. 7th Int. Conf. on Reliability, InfoCom Technologies and Optimization (ICRITO'2018); Aug. 29–31, 2018, Amity University, Noida, India*, 2018. For a preprint, see https://people.engr.ncsu.edu/brglez/publications.html.

[17] Franc Brglez. On Uncensored Global Stochastic Optimization in $\mathbb{R}^p/\mathbb{D}^p$ and the Efficiency of Multi-Walk Algorithms under Tableau Formulations. *Work in progress. For updates, see https://people.engr.ncsu.edu/brglez/publications.html*, 2018.

[18] Edward R. Tufte. *Beautiful Evidence*. Graphics Press LLC, 2006. https://www.edwardtufte.com/tufte/.

[19] Churchill Eisenhart. Realistic evaluation of the precision and accuracy of instrument calibration systems. *J. of Research of the NBS*, 67:161–187, 1963.

[20] Stanley D Rasberry. Accuracy in Analysis: The Role of Standard Reference Materials. *J. of Research of the NBS*, 93(3):213–216, 1988.

[21] A. Checchetti and A. Fantini. Experimental Determination of Planck's constant using Light Emitting Diodes (LEDs) and Photoelectric Effect. *World J. of Chemical Ed.*, 3(4):87–92, 2015.

[22] Katharine Mullen. Continuous Global Optimization in R. *Journal of Statistical Software, Articles*, 60(6):1–45, 2014.