9-2017

Sequential schemes for frequentist estimation of properties in statistical model checking

Cyrille JEGOUREL
1 Singapore University of Technology and Design

Jun SUN
Singapore Management University, junsun@smu.edu.sg

Jin Song DONG
Griffith University

Citation

JEGOUREL, Cyrille; SUN, Jun; and DONG, Jin Song. Sequential schemes for frequentist estimation of properties in statistical model checking. (2017). *Quantitative Evaluation of Systems: 14th International Conference, QEST 2017, Berlin, Germany, September 5-7: Proceedings*. 10503, 333-350.
Available at: https://ink.library.smu.edu.sg/sis_research/4715

Sequential Schemes for Frequentist Estimation of Properties in Statistical Model Checking*

Cyrille Jegourel[1], Jun Sun[1], and Jin Song Dong[2]

[1]Singapore University of Technology and Design, Singapore
[2]Griffith University, Australia
{cyrille.jegourel,sunjunhqq,dongjs1}@gmail.com

Abstract Statistical Model Checking (SMC) is an approximate verification method that overcomes the state space explosion problem for probabilistic systems by Monte Carlo simulations. Simulations might be however costly if many samples are required. It is thus necessary to implement efficient algorithms to reduce the sample size while preserving precision and accuracy. In the literature, some sequential schemes have been provided for the estimation of property occurrence based on predefined confidence and *absolute* or *relative* error. Nevertheless, these algorithms remain conservative and may result in huge sample sizes if the required precision standards are demanding. In this article, we compare some useful bounds and some sequential methods based on frequentist estimations. We propose outperforming and rigorous alternative schemes, based on Massart bounds and robust confidence intervals. Our theoretical and empirical analysis show that our proposal reduces the sample size while providing guarantees on error bounds.

1 Introduction

Probabilistic Model Checking (PMC) [16] is a formal verification method to analyse quantitative properties of probabilistic systems. PMC algorithms perform an exhaustive traversal of the state space of the system. However, real-world applications often involve multiple interacting components and the resulting state space becomes intractable. This limitation has led to the development of alternative methods like discrete event simulation and Statistical (Probabilistic) Model Checking (SMC) [23]. These simulation-based approaches require the use of an executable model of the system and then estimate the probability of a property based on simulations. SMC provides rigorous bounds of the error of the estimated results, based on robust statistical techniques (e.g., [19,4,21]). For real-world complex systems, SMC has a lot of potential as it requires little memory and remains very efficient for large systems. Finally, SMC is sometimes the only option for verifying many realistic models.

SMC also faces some specific problems. For example, simulations may be costly and time consuming. Moreover, the specifications of critical or important events are in practice tight. SMC must thus focus on additional statistical aspects to provide optimised sampling schemes while guaranteeing a rigorous confidence of the estimation. The

* **Acknowledgements.** Cyrille Jegourel and Jun Sun are partially supported by NRF grant RGNRF1501 and Jin Song Dong by the project: Reliable Prototyping Framework for Daily Living Assistance of Frail Ageing People (RELIANCE).

need of rigorous sampling schemes have been addressed from the early days in SMC [23,11] to more recent [10,8] just to cite a few. A key feature in designing a sampling procedure is to determine the number of simulations necessary to generate an estimation within acceptable margins of error and confidence. Bayesian SMC may be used to address this problem. However, in this approach, the probability to estimate must be given by a prior random variable whose density is based on previous experiments and knowledge about the system [25]. This limitation motivates the alternative frequentist estimation approaches. The scope of this article is restricted to this class of methods.

In [11], the authors discussed the notion of *absolute* and *relative* margin of error for SMC. To guarantee that the absolute error is bounded, they introduced a procedure relying on the Okamoto bound[1] that, given fixed confidence and error parameters, determines *a priori* the number of Bernoulli samples required, which is independent of the probability to estimate. Supporting relative errors (i.e., errors which depend on the probability to quantify) is more difficult, although theoretical bounds exist. Dagum et al. [7] proposed an *approximate* algorithm based on Bernstein's inequalities.

Approximate algorithms work by rough parameter estimations that are then reused in a stopping rule to update the number of simulations achieving the desired precision task. More recently, Watanabe proposed a sequential algorithm for bounding the relative error [22] based on a simpler stopping rule. The procedures described in [11] have been at least partially implemented in statistical model checkers like PRISM [16], PLASMA [14], APMC [12] and UPPAAL-SMC [8]. These sampling schemes are however very conservative notably when the probability to estimate is close to 0 or 1. Moreover, Dagum's algorithm was initially used to estimate the mean value of any random variable distributed in $[0, 1]$ and is thus not optimised for Bernoulli random variables.

In this article, our main goal is to provide better performing sampling schemes that *rigorously* fulfil absolute and relative error specifications. The key idea of our schemes is to define sequentially confidence intervals (CI) of the probability and then to apply Massart bounds, sharper than the Chernoff bounds, over the worst value of the CI to decide whether enough traces have been sampled. For this purpose, we also aim to clarify the two-sided "Chernoff" bounds for absolute and relative error specifications, to promote Massart bounds and last but not least, to give proofs of all these bounds. Indeed, the original theorems are one-sided and the two-sided versions must be clearly stated. The proofs are sometimes straightforward, at least for Theorems 3 and 5, but sometimes require more arguments. In particular, we could not find clear wordings and proofs of Theorem 2 and Theorem 4 in the literature. Finally, as far as we know, Theorems 6 and 7 are original as well as the algorithms using them. The proofs of the bounds can be found in the appendix.

In Section 2, we formally state the absolute and relative specifications which we want to fulfil. We also recall the basics of Monte Carlo estimation and some subtleties concerning coverage and CI. In Section 3, we introduce the Massart bounds. So far, they seem to suffer from a lack of recognition. For that reason, we present a comparison with the Chernoff bounds. We then describe some existing sampling schemes related to our problem in Section 4. In Section 5, we propose alternative sequential algorithms based on two inequalities, previously proven, which depend on the *coverage* of the probability.

[1] The Okamoto bound is sometimes called the Chernoff bound in the literature.

Finally, we show in Section 6 that these new schemes outperform the current approaches for the absolute and relative error problems by reducing significantly the sampling size. Section 7 concludes the article and leaves open questions for future work.

2 Background

In the following, a stochastic system $\mathcal{S}$ is interpreted as a set of interacting components in which the state is determined randomly with respect to a global probability distribution. Let $(\Omega, \mathcal{F}, \mu)$ be the probability space induced by the system with Ω a set of finite paths with respect to system's property ϕ, $\mathcal{F}$ a σ-algebra of Ω and μ the probability distribution defined over $\mathcal{F}$. Before going further, it is worth mentioning that SMC initially addressed the problem of verifying whether a property probability exceeds a threshold or not. This problem can be solved by using the sequential probability ratio test in hypothesis testing [23]. Other issues have been considered since, notably the estimation of the probability that a system property holds. In spite of similarities, both problems are different and in what follows, we focus on the estimation problem.

2.1 Statement of the problem

Given a probabilistic system $\mathcal{S}$, a property ϕ and a probability γ, we write $\mathcal{S} \models Pr(\phi) = \gamma$ if and only if the probability that a random execution of $\mathcal{S}$ satisfies ϕ is equal to γ. In principle, if γ is unknown, we can apply analytical methods to determine this value. However, due for example to numerical imprecisions, we often relax the constraints over γ and introduce the following notations:

$$\mathcal{S} \models^{a}_{\epsilon} Pr(\phi) = \gamma \quad \text{and} \quad \mathcal{S} \models^{r}_{\epsilon} Pr(\phi) = \gamma \tag{1}$$

The left formula means that a random execution of $\mathcal{S}$ satisfies ϕ with probability γ plus or minus an absolute error ϵ, i.e. $Pr(\phi) \in [\gamma - \epsilon, \gamma + \epsilon]$. The right formula means that a random execution of $\mathcal{S}$ satisfies ϕ with probability γ up to some relative error ϵ, i.e. $Pr(\phi) \in [(1-\epsilon)\gamma, (1+\epsilon)\gamma]$.

SMC applies on an executable system $\mathcal{S}$ and a property ϕ that is verified in finite time. In SMC, the satisfaction of property ϕ is quantified by a Bernoulli random variable of unknown mean γ. This mean is then approximated using a Monte Carlo estimation scheme. The output of the scheme is thus not an exact but an approximate value, given within certain error bounds and a confidence parameter δ that is the probability of outputting a false estimate. SMC thus requires a sampling scheme which outputs, after n samples, an estimate $\hat{\gamma}_n$ close to γ up to some absolute or relative ϵ-based error with probability greater or equal than $1 - \delta$. Formally, we write:

$$\mathcal{S} \models^{a}_{\epsilon,\delta} Pr(\phi) = \hat{\gamma}_n \quad \text{or} \quad \mathcal{S} \models^{r}_{\epsilon,\delta} Pr(\phi) = \hat{\gamma}_n \tag{2}$$

if and only if an algorithm outputs estimators while guaranteeing:

$$Pr(|\hat{\gamma}_n - \gamma| > \epsilon) \leq \delta \tag{3}$$

or respectively:

$$Pr(|\hat{\gamma}_n - \gamma| > \epsilon\gamma) \leq \delta. \tag{4}$$

We call (3) the absolute error specification and (4) the relative error specification. The goal of the article is thus to equip SMC with sampling algorithms that fulfil Specification (3) or (4) with as few samples as possible.

2.2 Monte Carlo Estimation

Let ω be a path sampled from space Ω with respect to distribution μ; z be a function from Ω to $\{0, 1\}$ assigning 1 if ω satisfies property ϕ and 0 otherwise and γ be the probability that an arbitrary path of the system satisfies ϕ. In SMC, the behaviour of function z is interpreted as a Bernoulli random variable Z with mean parameter γ. By definition, the average value γ is the integral of function z with respect to distribution μ over space Ω: $\gamma = \mathrm{E}_\mu[Z] = \int_\Omega z(\omega)\,\mathrm{d}\mu(\omega)$ and an estimator $\hat{\gamma}_n$ is given by the Monte Carlo method by drawing n independent samples $\omega_i \sim \mu$, $i \in \{1, \ldots, n\}$, as follows:

$$\hat{\gamma}_n = \frac{1}{n}\sum_{i=1}^{n} z(\omega_i) \approx \mathrm{E}_\mu[Z] \tag{5}$$

Let $m = \sum_{i=1}^{n} z(\omega_i)$ be the number of successes and $\sigma^2 = \gamma(1-\gamma)$ the variance of Z. In what follows, for sake of simplicity, we use both notations $\hat{\gamma}_n$ and m/n to denote the estimate.

Confidence Intervals and Coverage An estimator is given in general within a CI. However, in order to make use of the theorems presented in Section 5, we need to distinguish the notion of coverage and approximate CI.

Definition 1. *Given probability γ and a CI I, we call $C(\gamma, I) = Pr(\gamma \in I)$ the coverage of γ (by I).*

Denoting $\Phi(.)$ the standard normal distribution function and $z_{\delta/2} = \Phi^{-1}(1-\delta/2)$ the $(1-\delta/2)$th quantile of the normal distribution, the notional $(1-\delta)$-CI for γ is given by $I = \left[\hat{\gamma}_n - z_{\delta/2}\frac{\sigma}{\sqrt{n}}, \hat{\gamma}_n + z_{\delta/2}\frac{\sigma}{\sqrt{n}}\right]$ in virtue of the central limit theorem. However, in practice, σ^2 is replaced by a sample approximation $\hat{\sigma}_n^2 = \hat{\gamma}_n(1-\hat{\gamma}_n)/n$ (and if n is small, $z_{\delta/2}$ by $t_{\delta/2,n-1}$ the quantile of the Student's t-distribution with $n-1$ degrees of freedom). Then, an approximate $(1-\delta)$-CI $\tilde{I}$ is given by:

$$\tilde{I} = \left[\hat{\gamma}_n - z_{\delta/2}\hat{\sigma}_n, \hat{\gamma}_n + z_{\delta/2}\hat{\sigma}_n\right] \tag{6}$$

Unfortunately, the coverage of γ by an approximate CI $\tilde{I}$, may be significantly below the (desired) notional coverage: $C(\gamma, \tilde{I}) < C(\gamma, I) = 1-\delta$. More details about this topic are available in the appendix and in [2].

Exact Clopper-Pearson CI The algorithms proposed in Section 5 require an iterative computation of CI to evaluate a rigorous coverage of γ. For that purpose, we use the Clopper-Pearson $(1-\delta)$-CI [6]. This CI guarantees that the actual coverage is always equal to or above the nominal confidence level. In others words, a $(1-\delta)$-Clopper-Pearson CI J guarantees that $C(\gamma, J) \geq 1-\delta$ and its closed-form expression is easily computed: $J = [\,\beta^{-1}\left(\frac{\delta}{2}, m, n-m+1\right), \beta^{-1}(1-\frac{\delta}{2}, m+1, n-m)\,]$ with $\beta^{-1}(\delta, u, v)$ being the δ-th quantile of a Beta distribution parametrised by u and v.

Agresti-Coull CI As γ decreases, the Clopper-Pearson CI becomes more conservative. The Agresti-Coull CI consists in replacing the number of samples n by $n + z_\delta^2$ and the number of successes m by $m + z_\delta^2/2$ in the binomial CI (6). The CI is only approximate but still presents a good coverage close to the boundaries and may represent a good compromise between exactness and conservativeness (see [2] for more details).

3 Chernoff-Hoeffding-Okamato and Massart bounds

In the literature, the Chernoff bounds [4] refer to exponential decreasing bounds, in the number of simulations, of the probability of deviation between a Monte Carlo estimate and its mean. However, they exist under various forms, additive or multiplicative, one or two-sided, more or less "simplified". Moreover, tighter bounds have been established, notably in [18], but they still suffer from a lack of recognition. In this section, we intend to clear up confusion on the bounds by presenting a brief survey of the two-sided bounds and show the improvements achieved by the Massart bounds to give them the attention they deserve.

3.1 Absolute error bounds

Though the seminal work is due to Chernoff [4], the two-sided absolute error bound has been first stated for binomial distributions by M. Okamoto in [20].

Theorem 1 (Okamoto bound). *For any ϵ, $0 < \epsilon < 1$, we have the following inequality:*

$$Pr(|\hat{\gamma}_n - \gamma| > \epsilon) \leq 2\exp(-2n\epsilon^2) \tag{7}$$

Given ϵ, δ, writing out $\delta = 2\exp(-2n\epsilon^2)$, the Okamato bound can be used to determine a minimal number n of simulations to perform a Monte Carlo plan fulfilling the absolute error specification (3). The main advantage of the Okamoto bound is that it is independent of the value to estimate. However, the bound is very conservative and in many cases, a much lower sample size would achieve the same absolute error specification. Hoeffding provided a one-sided tighter exponential bound in [13]. We present below a two-sided version of his bound.

Theorem 2 (Absolute Error Hoeffding bound). *For any ϵ such that $0 < \epsilon < 1$ and γ such that $0 < \gamma < 1$, we have the following inequality:*

$$Pr(|\hat{\gamma}_n - \gamma| > \epsilon) \leq 2\exp\left(-n\epsilon^2 f(\gamma)\right) \tag{8}$$

where $f(\gamma) = \begin{cases} 1/(1-2\gamma)\,\log((1-\gamma)/\gamma) & \text{if } \gamma \neq 1/2 \\ 2 & \text{if } \gamma = 1/2 \end{cases}$

Surprisingly, we could not find a clear statement and a proof of this result in the literature. We thus present a proof in the appendix.

In this article, the Hoeffding bound is only presented because of its repute. Indeed, Massart established in [18] a sharper bound that holds if the absolute error ϵ is lower than probabilities γ and $1 - \gamma$. In what follows, we use the two-sided absolute and relative error versions of Massart bounds.

Theorem 3 (Absolute Error Massart bound). *For all γ such that $0 < \gamma < 1$ and any ϵ such that $0 < \epsilon < \min(\gamma, 1-\gamma)$, we have the following inequality:*

$$Pr(|\hat{\gamma}_n - \gamma| > \epsilon) \leq 2\exp\left(-n\epsilon^2 h_a(\gamma, \epsilon)\right) \tag{9}$$

$$\textit{where} \quad h_a(\gamma,\epsilon) = \begin{cases} 9/2\ ((3\gamma+\epsilon)(3(1-\gamma)-\epsilon))^{-1} & \textit{if } 0 < \gamma < 1/2 \\ 9/2\ ((3(1-\gamma)+\epsilon)(3\gamma+\epsilon))^{-1} & \textit{if } 1/2 \leq \gamma < 1 \end{cases}$$

3.2 Relative error bounds

In practice, the absolute error is set independently of γ. However, it could be that the approximation is meaningless, especially if the absolute error is large with respect to γ. In this case, setting a relative error that remains 'small' with respect of γ may be more adequate. The literature mentions a Chernoff-Hoeffding bound with relative error (e.g. [1]). This bound is known under multiple forms, more or less sharp and one or two-sided. For sake of consistency, we here provide a two-sided bound. As the existing literature adopts slightly different results, sometimes without providing their proof, we give a complete proof in the appendix adapted from two online references[2].

Theorem 4 (Relative Error Hoeffding bound). *For any ϵ, $0 < \epsilon < 1$ and γ, $0 < \gamma < 1$, we have the following inequality:*

$$Pr\left(|\hat{\gamma}_n - \gamma| > \epsilon\gamma\right) \leq 2\exp\left(-\frac{n\epsilon^2\gamma}{2+\epsilon}\right) \tag{10}$$

Finally, the Massart bound has a two-sided relative form.

Theorem 5 (Relative Error Massart bound). *For γ, $0 < \gamma < 1$ and any ϵ, $0 < \epsilon < (1-\gamma)/\gamma$, we have the following inequality:*

$$Pr(|\hat{\gamma}_n - \gamma| \geq \epsilon\gamma) \leq 2\exp\left(-n\epsilon^2 h_r(\gamma, \epsilon)\right) \tag{11}$$

$$\textit{with} \quad h_r(\gamma,\epsilon) = \begin{cases} 9\gamma/2\ ((3+\epsilon)(3-\gamma(3+\epsilon)))^{-1} & \textit{if } 0 < \gamma < 1/2 \\ 9\gamma/2\ ((3-\epsilon)(3-\gamma(3-\epsilon)))^{-1} & \textit{if } 1/2 \leq \gamma < 1 \end{cases}$$

Notional sample size If we let δ be equal to any of the right side expression of the inequalities given in Theorems (1) to (5), we can deduce a notional sample size n such that specification (3) or (4) is fulfilled. For example, using Theorem 5 given ϵ and δ, we only need to set $n > 1/(h_r(\gamma,\epsilon)\epsilon^2)\log(2/\delta)$ to satisfy the relative error specification (4). However, Hoeffding and Massart inequalities are not directly applicable because they depend on γ, in contrast to the Okamoto bound. But, they still have a theoretical interest: Figure 1 indicates for any notional γ the number of simulations necessary to produce an (ϵ, δ)-estimator according to the Okamoto, Hoeffding and Massart bounds. Though the bounds are approximately equivalent when γ is $1/2$, the bounds are far

[2] http://crypto.stanford.edu/∼blynn/pr/chernoff.html and www.cs.princeton.edu/courses/archive/fall09/cos521/Handouts/probabilityandcomputing.pdf

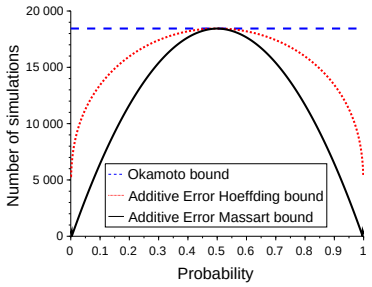


Figure 1: Okamoto (dash), Hoeffding (dot) and Massart (plain) bounds with absolute error $\epsilon = 0.01$ and confidence parameter $\delta = 0.05$.

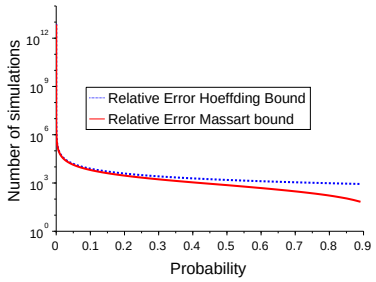


Figure 2: Hoeffding (dot) and Massart (plain) bounds with relative error $\epsilon = 0.1$ and confidence parameter $\delta = 0.05$.

apart when γ is away from $1/2$. Given $\epsilon = 0.01$, $\delta = 0.05$ and $\gamma = 0.05$ for example, the absolute error specification would be fulfilled with $n \geq 3283$ simulations according to the Massart bound instead of $n \geq 11276$ or $n \geq 18445$ for the respective Hoeffding and Okamoto bounds. Similarly for the relative error specification, Figure 2 shows that the Massart sample size is always lower than the Chernoff-Hoeffding sample size. The gain in sample size is more important when γ is high. With $\epsilon = 0.1$, $\delta = 0.05$, the ratio between Hoeffding and Massart sample sizes tends to decrease to 1.086 when γ tends to zero, that may still be non-negligible if sampling is time-costly.

4 Related work

In this section, we give a brief summary of existing sequential methods based on frequentis estimations to address specification (3) or (4). Some of them have already been implemented in SMC. We also recall that the specifications can be alternatively addressed by Bayesian SMC, not explored in this article, when beliefs and knowledge about the system are exploitable [25].

4.1 Schemes for the absolute error specification

Given ϵ and δ, the standard method to satisfy specification (3) is to compute a sample size n independently of probability γ using the Okamoto bound. Since there does not exist a bound independent of γ in the relative error case, the sequential schemes are mostly used to address specification (4) but they are not limited to it.

Simple scheme A simple idea could be to sample and update a $(1-\delta)$-CI until it is included into an interval $\hat{\gamma}_n \pm \epsilon$. This frequentist approach is implemented in UPPAAL-SMC [8]. However, though this technique may work more often if the CI are computed according to the Clopper-Pearson method, this scheme does not guarantee in general specification (3) for any δ, ϵ and γ (see for example [9,17]). For sake of understanding,

we added a brief but technical explanation in the appendix. It is however possible to pre-compute a value δ^* that guarantees a final coverage greater than $1-\delta$ (see [9]).

Chen's scheme [3] A promising sequential scheme which may work in practice, at least for some common values of ϵ and δ, is the work proposed by Chen in [3]. Chen's scheme also takes advantage of the Massart bounds. The idea is to sample while $n < 2\frac{\log(2/\delta)}{\epsilon^2}\left[1/4 - (|\hat{\gamma}_n - 1/2| - 2/3\epsilon)^2\right]$. Unfortunately, this rule only guarantees to produce an estimation which does not exceed the error bound ϵ on one side. So far, showing the other half of the bound has not been proven and was conjectured by the authors after some experiments.

4.2 Schemes for the relative error specification

In [11], the relative error specification is addressed by Dagum's algorithm.

Dagum's scheme [7] is a three-step procedure to perform an estimation of the mean of a general $[0,1]$-valued random variable X given relative error ϵ and confidence δ. The two first steps consist in providing a coarse estimation $\hat{\gamma}_k$ and a dispersion parameter $\hat{\rho}_l$. Finally, the third step provides the final estimation $\hat{\gamma}_n$ using $\hat{\gamma}_k$ and $\hat{\rho}_l$. The three steps are independent and depend on three different stopping rules, omitted here for sake of simplicity (see [7] for more details). The final sample size is thus given by $k+l+n$. Nevertheless, Dagum's scheme is based on coarser bounds than the Chernoff bounds. Moreover, this algorithm is used to estimate the mean of any random variable with support in $[0,1]$. Consequently, the scheme has a very general use but is not optimised for Bernoulli random variables.

Watanabe's scheme [22] In order to guarantee the relative error specification, Watanabe proposed to sample until the number of successes is greater than $\frac{3(1+\epsilon)}{\epsilon^2}\log\frac{2}{\delta}$. The main advantage is that this simple scheme does not require to perform pre-samples as in the first two steps of Dagum's algorithm. As far as we know, this scheme, more recent than Dagum's, has not been implemented in SMC.

5 A sequential scheme involving coverage

In this section, we present our sequential scheme for the absolute and relative error specification. Our scheme performs better than Watanabe and Dagum's scheme in the relative error case and, unlike the simple and Chen's schemes, is guaranteed to bound the error on both sides while strictly maintaining a coverage greater than $1-\delta$. Apart from the Okamoto bound, the inequalities presented in Section 3 require the knowledge of γ and they are thus not directly applicable. However, one may still exploit some information about probability γ. For example, depending on the problem, one may know or numerically evaluate with certainty a rough interval in which γ evolves. We present in the first subsection two theorems and the underlying sample sizes and, in the second subsection, our sampling schemes.

5.1 Bounds with coverage

The following theorems make use of the Massart bounds presented in Theorems 3 and 5 as they are sharper than the Chernoff-Hoeffding bounds.

Theorem 6 (Absolute Error Massart Bound with coverage). *Let a and b be the extrema of CI $I \in \mathcal{B}([0,1])$ and I^c be the complement of I in $[0,1]$:*

$$Pr\left(|\hat{\gamma}_n - \gamma| > \epsilon\right) \leq 2\exp\left(-n\epsilon^2 h_a(x,\epsilon)\right) + C(\gamma, I^c) \tag{12}$$

where function h_a is defined in Theorem 3 and $x = a$ if $b < 1/2$, $x = b$ if $a > 1/2$ and $x = 1/2$ if $1/2 \in I$.

By default, $a = 0$, $b = 1$, $C(\gamma, [0,1]^c) = 0$ and the theorem is consistent with the Okamoto bound. We remark that even if an accurate estimation of γ is not feasible to obtain within a reasonable time, Theorem 6 can exploit coarse but exact bounds a, b calculated analytically. In that case, we would have $C(\gamma, [a,b]^c) = 0$. Finally, a similar theorem involving relative error can be established.

Theorem 7 (Relative Error Massart Bound with coverage). *Let a be a (random) element of $[0,1]$ and h_r defined as in Theorem 5.*

$$Pr\left(|\hat{\gamma}_n - \gamma| > \epsilon\gamma\right) \leq 2\exp\left(-n\epsilon^2 h_r(a,\epsilon)\right) + C(\gamma, [0,a[) \tag{13}$$

Both theorems state that the probability of absolute or relative error is bounded by the respective Massart bound applied over the most pessimistic value of a CI plus the probability that the CI does not contain γ. We deduce from both theorems the following sample-size result:

Theorem 8. *Let $\delta' < \delta$ such that $C(\gamma, I^c) < \delta'$. (i) Under the conditions of Theorem 6, a Monte Carlo algorithm $\mathcal{A}$ that outputs an estimate $\hat{\gamma}_n$ fulfils Specification (3) if $n > \frac{1}{\min(h_a(a,\epsilon), h_a(b,\epsilon))\epsilon^2} \log \frac{2}{\delta - \delta'}$.*

(ii) Similarly, under the conditions of Theorem 7, a Monte Carlo algorithm $\mathcal{A}$ that outputs an estimate $\hat{\gamma}_n$ fulfils Specification (4) if $n > \frac{1}{h_r(a,\epsilon)\epsilon^2} \log \frac{2}{\delta - \delta'}$.

The proof is immediate in both cases once we set $\delta = 2s + \delta'$ with s being the respective exponential expressions of Theorems 6 or 7.

The bounds of Theorem 8 are more conservative than the bounds induced by Theorems 3 and 5 because the Massart bounds are evaluated in the most pessimistic value of CI $[a,b]$. In addition, our bound also takes into account the probability that γ is not in I, that implies an additional number of samples in the final sample size. In the absolute error case, if a CI I containing $1/2$ is determined, applying the previous theorem is unnecessary because the sample size is simply bounded with respect to the Okamoto bound. Similarly, if a (or b) is lower-bounded (or respectively upper-bounded) by $1/2$ but still close to $1/2$, the Okamoto bound is likely better. However, if γ is closer to 0 or 1, the logarithmic extra number of samples is largely compensated by the evaluation of the Massart bound in a or b.

5.2 Sequential algorithms

In the following, we present two new sampling schemes. Both of them require three inputs: an error parameter ϵ, and two confidence parameters δ and δ' such that $\delta' < \delta$. After each sample, we update a Monte Carlo estimator and a $(1-\delta')$-CI for γ. Then, the most pessimistic bound of the CI is used in the Massart function to compute a new minimal sample size n that satisfies Theorem 8. The process is repeated until the calculated sample size is lower than or equal to the current number of runs. We provide the pseudo-code of our Algorithms (1) and (2). Keywords GENERATE corresponds to a sample path generation and DETERMINE to the evaluation of the CI, slightly different in both schemes. Theorems 6 and 7 guarantee the correctness of our schemes since, for any tuple (m, n), if we are able to compute a $(1-\delta')$-CI I and its exact coverage, the deviation probability is bounded by δ defined as the sum of the coverage and the Massart function at n, ϵ and the most pessimistic value of I.

Absolute Error Sequential Algorithm We initiate the algorithm with a CI I_0 in which γ belongs (by default, $I_0 = [0, 1]$) and a worst-case (ϵ, δ)-sample size $n_0 = M$ with $M = \lceil \frac{1}{2\epsilon^2} \log \frac{2}{\delta} \rceil$ determined by the Okamoto bound ($\lceil . \rceil$ denotes the ceiling function). Once a trace $\omega^{(k)}$ is generated and monitored, the number of successes with respect to property ϕ and the total number of traces are updated. Then, an exact $(1-\delta')$-CI I_k is evaluated. Iteration after iteration, the CI width tends to shorten and becomes more and more accurate. Theorem 8-i is applied to determine a new sample size n_k, bounded from above by M if necessary. These steps are repeated until $k \geq n_k$ at which specification (3) is rigorously fulfilled.

Relative Error Sequential Algorithm We first assume the existence, in a practical case study, of a threshold γ_{min}, supposedly low, corresponding to a tolerated precision error (e.g. a floating-point approximation). Estimating a value below γ_{min} is then unnecessary. The maximal number of simulations is consequently bounded by $M = \lceil \frac{1}{\epsilon^2 h_r(\gamma_{min}, \epsilon)} \log \frac{2}{\delta} \rceil$. The relative error scheme is similar to the absolute error scheme. Note however that it is only necessary to determine a lower bound of I_k since h_r is a decreasing function in γ. Then, we determine a one-sided Clopper-Pearson $(1-\delta')$-CI of shape $[a_k, 1]$ with $a_k = \beta^{-1}(\delta', m, n-m+1)$. Theorem 8-ii is applied to determine a new sample size n_k, upper bounded by M if $a_k < \gamma_{min}$ and the steps are repeated until $k \geq n_k$. If the final output $\hat{\gamma}_k$ is higher than γ_{min}, Specification (4) is rigorously fulfilled. Otherwise, we can still output that γ is lower than γ_{min} with probability greater that $1-\delta$.

6 Experiment Results

Our methods significantly reduce the sampling size while rigorously guaranteeing the specifications when probability γ gets away from $1/2$ in the absolute error case and for any γ in the relative error case, in comparison to the methods that have been documented for SMC in [11]. Both methods can be easily used to improve existing SMC tools. To give a glimpse of their efficiency, we give the gain in sampling size obtained with our methods in Table 1 over 3 standard Prism benchmarks described in [24]: the tandem

Algorithm 1: Absolute Error Sequential Algorithm

Data:
ϵ, δ, δ' : the original parameters
$M = \lceil \frac{1}{2\epsilon^2} \log \frac{2}{\delta} \rceil$: the Okamoto bound
$k = 0$
$m = 0$: the number of successes
$n_k = M$
$I_k = [a_k, b_k]$ $[0, 1]$: the initial CI to which γ is known to belong

while $k < n_k$ **do**
  $k \leftarrow k + 1$
  GENERATE $\omega^{(k)}$
  $z(\omega^{(k)}) = \mathbb{1}(\omega^{(k)} \models \phi)$
  $m \leftarrow m + z(\omega^{(k)})$
  DETERMINE I_k
  if $1/2 \in I_k$ **then**
    $n_k = M$
  else if $b_k < 1/2$ **then**
    $n_k = \lceil \frac{2}{h_a(b_k,\epsilon)\epsilon^2} \log \frac{2}{\delta - \delta'} \rceil$
  else
    $n_k = \lceil \frac{2}{h_a(a_k,\epsilon)\epsilon^2} \log \frac{2}{\delta - \delta'} \rceil$
  $n_k \leftarrow \min(n_k, M)$

Output: $\hat{\gamma}_k = m/k$

Algorithm 2: Relative Error Sequential Algorithm

Data:
ϵ, δ, δ', γ_{min} : the original parameters
$M = \lceil \frac{1}{\epsilon^2 h_r(\gamma_{min},\epsilon)} \log \frac{2}{\delta} \rceil$
$k = 0$
$n_k = M$
$I_k = [a_k, 1] = [\gamma_{min}, 1]$: the initial CI in which γ is supposed to belong

while $k < n_k$ **do**
  $k \leftarrow k + 1$
  GENERATE $\omega^{(k)}$
  $z(\omega^{(k)}) = \mathbb{1}(\omega^{(k)} \models \phi)$
  $m \leftarrow m + z(\omega^{(k)})$
  DETERMINE I_k
  if $\gamma_{min} \geq a_k$ **then**
    $n_k = M$
  else
    $n_k = \lceil \frac{1}{\epsilon^2 h_r(a_k,\epsilon)} \log \frac{2}{\delta - \delta'} \rceil$
  $n_k \leftarrow \min(n_k, M)$

Output: $\hat{\gamma}_k = m/k$

queueing network in which queue capacities are equal to 3, the 10-station symmetric polling system and the 20-dependable workstation cluster. We respectively verify that,

	γ	APMC (ϵ, δ)	(AE) Gain	Dagum (ϵ, δ)	(RE) Gain
tandem	0.155132	$(0.01, 0.001)$	1.7	$(0.05, 0.001)$	5.18
polling	0.540786	$(0.001, 0.01)$	1	$(0.01, 0.01)$	3.65
cluster	5.160834×10^{-4}	$(10^{-4}, 0.05)$	399	$(0.2, 0.05)$	9

Table 1: Sampling size gains over standard Prism benchmarks

from their respective initial states, the system is full within 20 time units in the tandem example, that station 1 will be served before station 2 in the second example and that the QoS will drop below minimum of quality within 1000 time units in the third example. We refer to the appendix and [24] for more details concerning the models and the properties. In Prism, the Okamoto sampling size can be computed with the APMC method. For a given ϵ and δ, we report in column "(AE) Gain" the ratio between the Okamoto sampling size and our sampling size (average based on 5 experiments). For example, the property of the cluster model has probability $\gamma = 5.160834 \times 10^{-4}$ to occur. Given absolute error $\epsilon = 10^{-4}$ and confidence parameter $\delta = 0.05$, it requires 184443973 paths to guarantee Specification (3) when our method only requires 462077 paths to guarantee the same specification, which is 399 fewer samples. Similarly, given relative error ϵ and confidence parameters in column "Dagum (ϵ, δ)", "(RE) Gain" corresponds to the ratio of the 5-experiment average sampling sizes obtained by Dagum's algorithm and our method, necessary to fulfil Specification (4). The sampling sizes of these examples are given in the appendix.

Our methods are general and the class of probabilistic systems on which the sampling schemes can be applied does not really matter as long as the systems are executable and the executions can be monitored. In what follows, we evaluate our sampling schemes on a small benchmark, available in the appendix, that can be easily investigated using model checker Prism [16] to corroborate our results.

6.1 Absolute Error Scheme Results

We compare our algorithm with the simple and Chen's schemes. To guarantee specification (3) in the simple scheme, one can use the algorithm proposed by Frey in [9]. This procedure pre-computes a value δ^* that guarantees a final coverage greater than $1 - \delta$ when the CI are computed according to the Clopper-Pearson method. For each couple of successes and trials (m, n) where n is smaller than the Okamoto bound M, the algorithm computes the number of sequences of observations $h(m, n, \epsilon)$ that lead to the output m/n. Unfortunately, we were unable to get results for ϵ smaller than 0.1 due to overflows of values $h(m, n, \epsilon) > 10^{309}$ in addition to an excessive amount of time required by this recursive computation. Thus, we used the default $\delta^* = \delta$.

We repeated each set of experiments 200 times with the three schemes for several values of γ, ϵ and δ. We estimated the empirical coverage by the number of times the specification (3) is fulfilled divided by 200 and computed the average, the standard deviation and the extrema values of the sample size and of the estimations $\hat{\gamma}$. For sake of clarity, as our results are consistent for all ϵ, δ and are symmetric with respect to $\gamma = 1/2$, we summarize the most relevant results for $\epsilon = 0.01$, $\delta = 0.05$ and $0 < \gamma \leq 1/2$ in Table 2. More details are provided for every scheme and set of experiments in the appendix. For every ϵ and δ, the sampling size is significantly lower for the

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage (simple)	1	0.965	**0.94**	0.96	0.965	0.975	**0.945**
$\hat{\gamma}$ min (simple)	0	0	**0.007**	**0.036**	**0.087**	0.288	**0.484**
$\hat{\gamma}$ max (simple)	0.013	0.021	0.029	0.062	**0.113**	**0.316**	**0.513**
N mean (simple)	518	729	1107	2172	3777	8278	9703
Coverage (Chen)	1	0.98	1	0.995	1	0.995	0.995
$\hat{\gamma}$ min (Chen)	0	0	0.011	0.04	0.091	0.292	0.492
$\hat{\gamma}$ max (Chen)	0.01	0.017	0.028	0.059	0.107	0.31	0.511
N mean (Chen)	810	1171	1900	3946	7035	15684	18444
Coverage (new)	1	0.99	995	0.995	0.995	1	1
$\hat{\gamma}$ min (new)	0	0	0.01	0.039	0.089	0.291	0.491
$\hat{\gamma}$ max (new)	0.011	0.019	0.027	0.059	0.106	0.309	0.51
N mean (new)	831	1229	2064	4474	8161	18434	18445

Table 2: Results of the Absolute Error scheme with $\epsilon = 0.01$ and $\delta = 0.05$

simple scheme than for Chen and our schemes. However, the empirical coverage is below $1 - \delta$ for some γ (in bold and red in the table). For example, Table 2) indicates an empirical coverage of 0.94 for $\epsilon = 0.01$, $\delta = 0.05$, and $\gamma = 0.02$. Moreover, we remark that for every set of experiments, the simple scheme outputs at least one estimation that exceeds $\gamma \pm 1.25\epsilon$ (in bold and red in the table). This indicates that the difference between the estimation and γ exceeds the absolute error ϵ by more than 25%, that may consequently lead to important analysis errors. In comparison, the difference between $\hat{\gamma}$ and γ never exceeds ϵ by more than 10% in both other schemes. We thus do not recommend to use the simple scheme if specification (3) is rigorously prescribed. The theoretical expectations of Chen and our schemes are empirically confirmed: the coverage is significantly above $1 - \delta$ in each case (> 0.95 in Table 2). Specification (3) is thus strictly satisfied. Chen's scheme shows a slightly better performance than our algorithm in terms of sampling size. However, we recall that Chen only guarantees that the estimation does not exceed the error bound ϵ on one side. For that reason, we recommend to use our algorithm that seems to be reasonably more conservative.

Figure 3a shows an empirical plot of the sample size as a function of probability γ. In this experiment, we let the sample size be greater than the Okamoto bound (dotted blue line) to illustrate the gap between the empirical and the notional bounds. With the sampling algorithm 1 described in Section 5, the sample size would be bounded in virtue of Okamoto's inequality between 0.3 and 0.7. Note that the empirical plot has no particular meaning but is a guide to the eye that illustrates the behaviour of our algorithm. As expected, the gain is larger close to 0 and 1. For $\gamma = 0.02$, the Okamoto sample size (18445) is divided in average by 9. The empirical sample size is always maintained above the notional Massart sample size, indicating that the sample size has not been mistakenly minimised due to a wrong CI.

6.2 Relative Error Scheme Results

We repeated 200 times Dagum's, Watanabe's and our relative error schemes for eight values of γ with several ϵ and δ. We reported the results for several ϵ and δ in Table 3.

γ	0.9	0.7	0.5	0.3	0.1	0.05	0.01	0.001
N mean Dagum, $(\epsilon, \delta) = (0.1, 0.01)$	1871	4402	9056	19703	74064	152757	803572	8124356
N mean Dagum, $(\epsilon, \delta) = (0.1, 0.05)$	1360	3160	6412	14253	52432	111703	570763	5787456
N mean Dagum, $(\epsilon, \delta) = (0.05, 0.05)$	3162	8912	19244	43276	163084	346269	1800585	18208080
N mean Dagum, $(\epsilon, \delta) = (0.05, 0.01)$	4394	12337	26677	60263	226889	479164	2467430	25300472
N mean W., $(\epsilon, \delta) = (0.1, 0.01)$	1942	2498	3501	5836	17479	35006	175092	1746713
N mean W., $(\epsilon, \delta) = (0.1, 0.05)$	1353	1738	2439	4048	12207	24362	122029	1218779
N mean W., $(\epsilon, \delta) = (0.05, 0.05)$	5163	6634	9299	15453	46496	92950	465144	4650289
N mean W., $(\epsilon, \delta) = (0.05, 0.01)$	7416	9540	13347	22235	66756	133581	665536	6677525
N mean New, $(\epsilon, \delta) = (0.1, 0.01)$	202	623	1373	3043	11365	23812	122426	1236491
N mean New, $(\epsilon, \delta) = (0.1, 0.05)$	137	441	991	2204	8208	17356	88838	895496
N mean New, $(\epsilon, \delta) = (0.05, 0.05)$	476	1631	3737	8473	32175	67850	348706	3515688
N mean New, $(\epsilon, \delta) = (0.05, 0.01)$	669	2266	5151	11675	44346	93236	482998	4871059

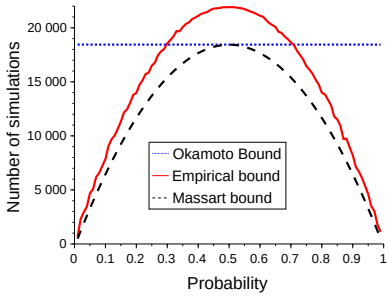
Table 3: Sample size average of the Relative Error schemes, given ϵ and δ.

More detailed tables, containing the descriptive statistics of the sample sizes, are available in the appendix. The average sample sizes are drawn for each scheme on Figure 3c. We did not report the coverage of the sampling schemes because specification (4) was largely satisfied in the three cases. However, Dagum scheme is very conservative in sample size as Table 3 and Figure 3c illustrate. We observe that our scheme is better than Watanabe's for all values of γ, especially when γ tends to 1. As γ decreases, the Clopper-Pearson CI becomes more conservative but our algorithm still presents better performances. However, when γ is below 0.05, the conservativeness of the Clopper-Pearson CI becomes too significant and exponentially impacts the sample size. Once the number of simulations k exceeds 1000 and $\hat{\gamma}_k \in [1/k, 0.04]$, we thus replaced the evaluation of the Clopper-Pearson CI by the Agresti-Coull CI. The results in the last two columns of Table 3 are obtained using the Agresti-Coull CI. The approximation maintains the highest performance and seems to be a good alternative to exact CI. A deeper investigation is left to future work, but even if the lower bound of the exact CI is below the lower bound of the Agresti-Coull CI, their difference is likely tight and reusing a slightly too optimistic value in the Massart bound is unlikely to pose problem.

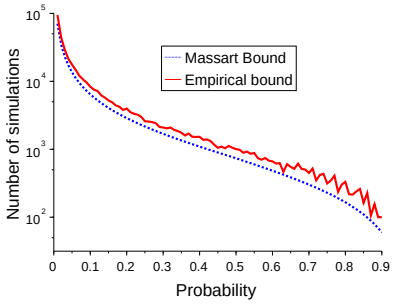
We compare in Figure 3b the empirical plot of the sample size as a function of probability γ with the notional bound. As for the absolute error case, the empirical plot is always maintained above the notional Massart sample size. Figure 3d shows the typical evolution of the CI bounds for the absolute and relative error problem with respectively, $\gamma = 0.05$ and $\gamma = 0.1$ and shows their accuracy and reliability over time.

7 Conclusion

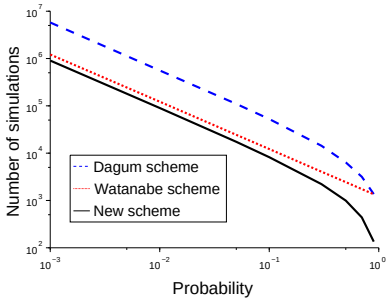
The focus of this paper was to minimise using sequential schemes based on frequentist estimations the sampling size necessary to estimate a property with absolute or relative error in comparison to the standard methods in SMC. To build estimators that fulfil Specifications (3) or (4), we proved two inequalities and presented two sequential algorithms based on Massart bounds and coverage of probability γ. The comparison with the schemes commonly used in SMC showed significant improvements. We leave for future work theoretical improvements of the algorithms, notably concerning the choice



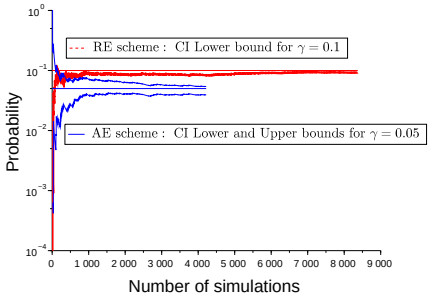
(a) Notional Okamoto (dot) and Massart (dash) bounds versus empirical results (absolute error $\epsilon = 0.01$ and confidence parameter $\delta = 0.05$).



(b) Massart bounds (dot) versus empirical bounds (relative error $\epsilon = 0.1$ and confidence parameter $\delta = 0.05$).



(c) Comparison between Dagum-and-al. (above), Watanabe (middle) and new (below) relative error schemes ($\epsilon = 0.1$ and $\delta = 0.05$).



(d) Evolution of the CI bounds for the Absolute and Relative Error schemes, respectively in blue and red with $\gamma = 0.05$ and 0.1.

Figure 3: Experimental results

of δ'. Finally, it is worth recalling that all the Monte Carlo sampling schemes anyway require a lot of samples for rare event estimation. Though the problem of designing sampling schemes for Binomial estimators is well-documented, the lack of exact concentration inequalities for importance sampling [5] and splitting estimators [15] in SMC makes the design of robust sampling procedures challenging in the rare event context.

References

1. D. Angluin and L. Valiant. Fast Probabilistic Algorithms for Hamiltonian Circuits and Matchings. *J. Comput. Syst. Sci.*, 18(2):155–193, 1979.
2. L. Brown, T. Cai, and A. DasGupta. Interval Estimation for a Binomial Proportion. *Statistical Science*, 16(2):101–133, 2001.
3. Jianhua Chen. Properties of a New Adaptive Sampling Method with Applications to Scalable Learning. In *WI, Atlanta*, pages 9–15, 2013.

4. H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23(4):493–507, 1952.
5. E. Clarke and P. Zuliani. Statistical Model Checking for Cyber-Physical Systems. In *9th International Symposium, ATVA, Taipei*, pages 1–12, 2011.
6. C.J. Clopper and E.S. Pearson. The Use of Confidence or Fiducial Limits illustrated in the Case of the Binomial. *Biometrika*, 26:404–413, 1934.
7. P. Dagum, R.M. Karp, M.L., and S.M. Ross. An Optimal Algorithm for Monte Carlo Estimation. *SIAM J. Comput.*, 29(5):1484–1496, 2000.
8. A. David, K.G. Larsen, A. Legay, M. Mikucionis, and D.B. Poulsen. Uppaal SMC tutorial. *STTT*, 17(4):397–415, 2015.
9. J. Frey. Fixed-Width Sequential Confidence Intervals for a Proportion. *The American Statistician*, 64(3):242–249, 2010.
10. R. Grosu, D. Peled, C.R. Ramakrishnan, S.A. Smolka, S.D. Stoller, and J. Yang. Using statistical model checking for measuring systems. In *6th International Symposium, ISoLA, Corfu*, pages 223–238, 2014.
11. T. Hérault, R. Lassaigne, F. Magniette, and S. Peyronnet. Approximate probabilistic model checking. In Bernhard Steffen and Giorgio Levi, editors, *VMCAI*, volume 2937 of *LNCS*, pages 307–329. Springer, 2004.
12. T. Hérault, R. Lassaigne, and S. Peyronnet. APMC 3.0: Approximate Verification of Discrete and Continuous Time Markov Chains. In *3rd Intl Conf. QEST, Riverside*, pages 129–130, 2006.
13. Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.
14. C. Jégourel, A. Legay, and S. Sedwards. A Platform for High Performance Statistical Model Checking - PLASMA. In *18th Intl Conference, TACAS 2012, Tallinn, 2012*, pages 498–503, 2012.
15. C. Jegourel, A. Legay, and S. Sedwards. Importance Splitting for Statistical Model Checking Rare Properties. In *CAV, Saint-Petersburg*, volume 8044 of *LNCS*, pages 576–591, 2013.
16. Marta Z. Kwiatkowska, Gethin Norman, and David Parker. PRISM 2.0: A Tool for Probabilistic Model Checking. In *QEST*, pages 322–323. IEEE, 2004.
17. W. Liu and B.J.R. Bailey. Sample size determination for constructing a constant width confidence interval for a binomial success probability. *Statistics and Probability Letters*, (56):1–5, 2002.
18. P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18:1269–1283, 1990.
19. N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–341, September 1949.
20. Masashi Okamoto. Some Inequalities Relating to the Partial Sum of Binomial Probabilities. *Annals of the Institute of Statistical Mathematics*, 10:29–35, 1958.
21. Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.
22. Osamu Watanabe. Sequential Sampling Techniques for Algorithmic Learning Theory. *Theoretical Computer Science*, 348:3–14, 2005.
23. H. Younes. *Verification and Planning for Stochastic Processes with Asynchronous Events*. PhD thesis, Carnegie Mellon University, 2004.
24. Håkan L. S. Younes, Marta Z. Kwiatkowska, Gethin Norman, and David Parker. Numerical vs. statistical probabilistic model checking: An empirical study. In *TACAS 2004, Barcelona, Spain*, pages 46–60, 2004.
25. P. Zuliani, A. Platzer, and E. M. Clarke. Bayesian Statistical Model Checking with Application to Stateflow/Simulink Verification. *Formal Methods in System Design*, 43(2):338–367, 2013.

Appendix A Content of the Appendices

In what follows, we provide in Appendix B the proofs of the theorems presented in the article. Due to the page limit and for sake of simplicity, some technicalities are omitted in the article. For these reasons, we briefly answer some common questions that readers may have about approximate confidence intervals and the simple scheme in Appendix C. We however refer to the references mentioned in the article for more details. In Appendix D, we describe the group repair model that we used in Section 6. We notably provide the Prism code of the model and the property and a table of correspondence between failure parameter α and γ. We also provide the settings of the other models (tandem, polling and cluster) for the readers. In Appendix E, we provide more results with different ϵ and δ for various γ.

Appendix B Proofs

In what follows, for more convenience and readability, we use both e^x and $\exp(x)$ to denote the exponential of a real number x.

B.1 Proof of Theorem 2 [Absolute Error Hoeffding bound]

Proof. Let us first introduce a few notations. We define f as the following function:

$$\begin{array}{rcl} f :]0,1[& \longrightarrow & \mathbb{R} \\ \gamma & \longmapsto & \frac{1}{1-2\gamma}\log\left(\frac{1-\gamma}{\gamma}\right) \end{array}$$

Let g be the function defined as:

$$\begin{array}{rcl} g :]0,1[& \longrightarrow & \mathbb{R} \\ \gamma & \longmapsto & \frac{1}{2\gamma(1-\gamma)} \end{array}$$

Note that f is defined by continuity in $1/2$ and that $f(1/2) = 2$ and that $g(\gamma) > 0$. Let us recall the result established by Hoeffding in [13]:

Theorem 9 (Hoeffding's bound). *For any ϵ, $0 < \epsilon < 1$, we have the following inequalities:*

$$\forall 0 < \gamma < \frac{1}{2}, \quad Pr(\hat{\gamma}_n - \gamma > \epsilon) \leq \exp\left(-n\epsilon^2 f(\gamma)\right) \tag{14}$$

and

$$\forall \frac{1}{2} \leq \gamma < 1, \quad Pr(\hat{\gamma}_n - \gamma > \epsilon) \leq \exp\left(-n\epsilon^2 g(\gamma)\right) \tag{15}$$

In order to prove the two-sided version presented in the article, we need to prove the following lemma.

Lemma 1. $\forall \gamma \in]0,1[\,, \quad f(\gamma) \leq g(\gamma)$.

Proof. Proving this lemma is equivalent to prove that $h(\gamma) \leq 0$ with

$$h(\gamma) = |1-2\gamma|\frac{f(\gamma)}{g(\gamma)} - |1-2\gamma|.$$

The derivative h' of h is equal to $\mathbf{sgn}(1-2\gamma) \times 2(1-2\gamma)\log\left(\frac{1-\gamma}{\gamma}\right)$ with $\mathbf{sgn}(.)$ the function that assigns to an element x of $\mathbb{R}$ its sign (1 if $x \geq 0$, -1 otherwise). From the sign of h', we deduce the variation of h (increasing between 0 and $1/2$ and decreasing between $1/2$ and 1). The maximum is thus reached in $1/2$ and $h(1/2) = 0$. □

Let $\mu = 1-\gamma$ and $\hat{\mu}_n = 1-\hat{\gamma}_n$.

$$\begin{aligned} Pr(|\hat{\gamma}_n - \gamma| > \epsilon) &= Pr(\hat{\gamma}_n - \gamma > \epsilon) + Pr(\hat{\gamma}_n - \gamma < -\epsilon) \\ &= Pr(\hat{\gamma}_n - \gamma > \epsilon) + Pr(\hat{\mu}_n - \mu > \epsilon) \end{aligned}$$

If $0 < \gamma < \frac{1}{2}$, as $\mu = 1-\gamma$, we use twice the Hoeffding's bound to write:

$$Pr(|\hat{\gamma}_n - \gamma| > \epsilon) \leq \exp\left(-n\epsilon^2 f(\gamma)\right) + \exp\left(-n\epsilon^2 g(\mu)\right)$$

And if $\frac{1}{2} \leq \gamma <$, similarly, we write:

$$Pr(|\hat{\gamma}_n - \gamma| > \epsilon) \leq \exp\left(-n\epsilon^2 g(\gamma)\right) + \exp\left(-n\epsilon^2 f(\mu)\right)$$

However, for all $\gamma \in]0,1[$, $f(\gamma) = f(1-\gamma)$ and $g(\gamma) = g(1-\gamma)$. Thus, both previous inequations may be simplified as follows:

$$\begin{aligned} Pr(|\hat{\gamma}_n - \gamma| > \epsilon) &\leq \exp\left(-n\epsilon^2 f(\gamma)\right) + \exp\left(-n\epsilon^2 g(\gamma)\right) \\ &\leq 2\exp\left(-n\epsilon^2 \min(f(\gamma), g(\gamma)\right) \end{aligned}$$

The proof is achieved by the use of Lemma 1. □

B.2 Proof of Theorem 3 [Absolute Error Massart bound]

Proof. According to Massart's seminal work [18], we already have the following inequality for γ, $0 < \gamma < 1$ and any ϵ, $0 < \epsilon < \min(\gamma, 1-\gamma)$:

$$Pr(\hat{\gamma}_n - \gamma > \epsilon) \leq \exp\left(-\frac{9n\epsilon^2}{2(3\gamma+\epsilon)(3(1-\gamma)-\epsilon)}\right) \tag{16}$$

Then, setting $\mu = 1-\gamma$ and $\hat{\mu}_n = 1-\hat{\gamma}_n$, dually, we have:

$$Pr(\hat{\mu}_n - \mu > \epsilon) \leq \exp\left(-\frac{9n\epsilon^2}{2(3\mu+\epsilon)(3(1-\mu)-\epsilon)}\right) \tag{17}$$

Rewriting the expression with respect to γ, we end up with the two-sided bound:

$$\begin{aligned} Pr(|\hat{\gamma}_n - \gamma| > \epsilon) &\leq e^{-\frac{9n\epsilon^2}{2(3\gamma+\epsilon)(3(1-\gamma)-\epsilon)}} + e^{-\frac{9n\epsilon^2}{2(3(1-\gamma)+\epsilon)(3\gamma-\epsilon)}} \\ &\leq 2\exp\left(-\frac{9n\epsilon^2}{2}\min\left(\frac{1}{(3\gamma+\epsilon)(3(1-\gamma)-\epsilon)}, \frac{1}{(3(1-\gamma)+\epsilon)(3\gamma-\epsilon)}\right)\right) \end{aligned}$$

The right-side expression is not very convenient to manipulate. We thus evaluate which exponential of the sum dominates the other. Let $A = (3\gamma + \epsilon)(3(1-\gamma) - \epsilon)$ and $B = (3(1-\gamma) + \epsilon)(3\gamma - \epsilon)$. After simplification, we get:

$$A - B = 6\epsilon(1 - 2\gamma)$$

that is greater than 0 if and only if $0 < \gamma \leq 1/2$. The conclusion follows immediately. □

B.3 Proof of Theorem 4 [Relative Error Hoeffding bound]

The following proof has been adapted from the partial proofs available online[3].

Proof. Let $S_n = \sum_{i=1}^{n} z_i$ be the sum of the independent Bernoulli observations z_i. To recall, by definition, $\hat{\gamma}_n = S_n/n$. Let us first bound $Pr\left(\hat{\gamma}_n - \gamma > \epsilon\gamma\right)$ or equivalently, $Pr\left(S_n > (1+\epsilon)n\gamma\right)$. For all $t > 0$, we have:

$$Pr\left(S_n > (1+\epsilon)n\gamma\right) = Pr\left(e^{tS_n} > e^{t(1+\epsilon)n\gamma}\right)$$

Then, by the Markov's inequality,

$$Pr\left(S_n > (1+\epsilon)n\gamma\right) \leq \frac{E\left[e^{tS_n}\right]}{e^{t(1+\epsilon)n\gamma}}$$

Then since the z_i are independent:

$$E\left[e^{tS_n}\right] = E\left[e^{t\sum_{i=1}^{n} z_i}\right] = E\left[\prod_{i=1}^{n} e^{tz_i}\right] = \prod_{i=1}^{n} E\left[e^{tz_i}\right]$$

z_i are Bernoulli random variables of parameter γ. We can thus evaluate $E\left[e^{tz_i}\right]$ easily. With probability γ, $e^{tz_i} = e^t$ and with probability $1-\gamma$, $e^{tz_i} = 1$. So, $E\left[e^{tz_i}\right] = 1 + \gamma(e^t - 1)$. Moreover, for all $x > 0$, $1 + x < e^x$. Thus,

$$\prod_{i=1}^{n} E\left[e^{tz_i}\right] = (1 + \gamma(e^t - 1))^n < e^{n\gamma(e^t - 1)}$$

Hence,

$$Pr\left(\hat{\gamma}_n - \gamma > \epsilon\gamma\right) \leq \frac{e^{n\gamma(e^t-1)}}{e^{t(1+\epsilon)n\gamma}} \tag{18}$$

This inequation is valid for all $t > 0$, in particular for the value t that minimises the right-hand side. By differentiation, we can show that the minimum is reached when $t = \log(1+\epsilon)$. Finally, after rewrital, we end up with a nice multiplicative Chernoff bound:

$$\begin{aligned} Pr\left(\hat{\gamma}_n - \gamma > \epsilon\gamma\right) &\leq \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{n\gamma} \\ &\leq e^{n\gamma(\epsilon - (1+\epsilon)\log(1+\epsilon))} \end{aligned}$$

[3] http://crypto.stanford.edu/∼blynn/pr/chernoff.html and www.cs.princeton.edu/courses/archive/fall09/cos521/Handouts/probabilityandcomputing.pdf

Lemma 2. $\forall x \geq 0$, $\log(1+x) \geq \frac{2x}{2+x}$.

Proof. Let h the function defined on $\mathbb{R}$ as $h(x) = \log(1+x) - \frac{2x}{2+x}$. The derivative h' is:

$$h'(x) = \frac{1}{1+x} - \frac{4}{(2+x)^2} = \frac{x^2}{(1+x)(2+x)^2} \geq 0. \tag{19}$$

So h is an increasing function, thus its global minimum is reached in 0. But $h(0) = 0$. □

Using Lemma 2, we obtain:

$$Pr\left(\hat{\gamma}_n - \gamma > \epsilon\gamma\right) \leq \exp\left(n\gamma\left(\epsilon - (1+\epsilon)\frac{2\epsilon}{2+\epsilon}\right)\right) \tag{20}$$

$$\leq \exp\left(-\frac{n\epsilon^2\gamma}{2+\epsilon}\right) \tag{21}$$

For the lower tail $P(\gamma - \hat{\gamma}_n > \epsilon\gamma)$, we proceed as before. We apply Markov's inequality, then use the approximation $1 + x < e^x$, and finally choose t to minimize the bound, that is $t = \log(1-\epsilon)$, to obtain:

$$Pr\left(\gamma - \hat{\gamma}_n > \epsilon\gamma\right) \leq \left(\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right)^{n\gamma} \tag{22}$$

$$\leq e^{n\gamma(-\epsilon-(1-\epsilon)\log(1-\epsilon))} \tag{23}$$

Recall that for $x < 1$, the Taylor expansion of $-\log(1-x)$ is $\sum_{k=1}^{\infty} \frac{x^k}{k}$. Thus, as $\epsilon < 1$, $\log(1-\epsilon) > -\epsilon + \epsilon^2/2$. Substituting this inequality in Expression (22) leads, after simplification, to:

$$Pr\left(\hat{\gamma}_n - \gamma > \epsilon\gamma\right) \leq \exp\left(-\frac{n\epsilon^2\gamma}{2}\right) \tag{24}$$

Finally, we have:

$$\begin{aligned} Pr\left(|\hat{\gamma}_n - \gamma| > \epsilon\gamma\right) &\leq Pr\left(\hat{\gamma}_n - \gamma > \epsilon\gamma\right) + Pr\left(\gamma - \hat{\gamma}_n > \epsilon\gamma\right) \\ &\leq \exp\left(-\frac{n\epsilon^2\gamma}{2+\epsilon}\right) + \exp\left(-\frac{n\epsilon^2\gamma}{2}\right) \end{aligned}$$

The proof of the Theorem is achieved noticing that $1/2 \geq 1/(2+\epsilon)$. □

B.4 Proof of Theorem 5 [Relative Error Massart bound]

Theorem 5 is just a particular case of Theorem 3. Let $0 < \epsilon < \min(1, \frac{1-\gamma}{\gamma})$ and $\epsilon' = \epsilon\gamma$. By construction, $0 < \epsilon' < \min(\gamma, 1-\gamma)$. Thus, we can apply Theorem 3 and we get:

$$Pr(|\hat{\gamma}_n - \gamma| > \epsilon') \leq 2\exp\left(-n\epsilon'^2 h_a(\gamma, \epsilon')\right) \tag{25}$$

The theorem is straightforward after the replacement of ϵ' by $\epsilon\gamma$ and simplification. □

B.5 Proof of Theorem 6 [Absolute Error Massart Bound involving knowledge]

Proof. First, let us establish the following lemma:

Lemma 3. *For any events $A, B \in \mathcal{F}$, denoting B^c the complement of B, we have:*

$$Pr(A) \leq Pr(A \mid B) + Pr(B^c)$$

.

Proof. $A = (A \cap B) \sqcup (A \cap B^c)$ with $\sqcup$ denoting a disjoint union. So, $Pr(A) = Pr(A \cap B) + Pr(A \cap B^c)$. But, by the Bayes Theorem, $Pr(A \mid B)Pr(B) = Pr(A \cap B)$ and $A \cap B^c \subset B^c$, thus $Pr(A \cap B) \leq Pr(A \mid B)$ and $Pr(A \cap B^c) \leq Pr(B^c)$. □

Then, by applying Lemma 3 on events $A = \{|\hat{\gamma}_n - \gamma| \geq \epsilon\}$ and $B = \{\gamma \in I\}$, we get:

$$Pr\left(|\hat{\gamma}_n - \gamma| > \epsilon\right) \leq Pr\left(|\hat{\gamma}_n - \gamma| > \epsilon \mid \gamma \in I\right) + Pr\left(\gamma \notin I\right) \tag{26}$$

First, by definition, $Pr\left(\gamma \notin I\right) = C(\gamma, I^c)$. Then, under hypothesis $a \leq \gamma \leq b$, the absolute error bounds imply that the first probability of the right expression is bounded by:

$$Pr\left(|\hat{\gamma}_n - \gamma| > \epsilon \mid \gamma \in I\right) \leq 2\exp\left(-n\epsilon^2 \min_{\gamma \in I} f(\gamma)\right)$$

A simple function analysis shows that f strictly decreases between 0 and $1/2$ and strictly increases between $1/2$ and 1. Its maximum is thus reached at one of the bounds of interval $[a, b]$, a if $a \geq 1/2$ and b if $b \leq 1/2$. □

B.6 Proof of Theorem 7 [Relative Error Massart Bound involving knowledge]

Proof. Let $I = [a, 1]$. We apply Lemma 3 on events $A = \{|\hat{\gamma}_n - \gamma| > \epsilon\gamma\}$ and $B = \{\gamma \notin I\}$. We get:

$$\begin{aligned} Pr\left(|\hat{\gamma}_n - \gamma| > \epsilon\gamma\right) &\leq Pr\left(|\hat{\gamma}_n - \gamma| > \epsilon\gamma \mid \{\gamma \in I\}\right) + Pr\left(\gamma < a\right) \\ &\leq 2\exp\left(-n\epsilon^2 \min_{\gamma \in I} \zeta_R(\gamma)\right) + Pr\left(\gamma < a\right) \end{aligned}$$

Similarly, function analysis shows that h increases on $]0, 1]$. The minimum is thus reached in a. □

Appendix C Questions

C.1 What are the problems with approximate confidence intervals?

First of all, the standard confidence interval results from the central limit theorem. This theorem is an asymptotic result involving the normal distribution. Thus, this asymptotic theorem should not be used if the central limit approximation is not accurate. But the heuristics provided by statistical textbooks are often imprecise (e.g. "the number of samples n must be large") or unverifiable (e.g. " $n\gamma(1-\gamma)$ must be greater than 5") as γ is unknown. Last but not least, even if the qualifications of the central limit theorem are true, the coverage of γ by an approximate confidence interval $\tilde{I}$, may be significantly below the (desired) notional coverage: $C(\gamma, \tilde{I}) < C(\gamma, I) = 1 - \delta$. We refer the readers to [2] for the excellent discussion about the coverage of confidence intervals.

γ	0.001	0.005	0.01	0.02	0.05	0.1	0.3	0.5	0.7	0.9
α	0.2384	0.2784	0.2978	0.3186	0.3491	0.3755	0.4303	0.4723	0.5272	0.7325

Table 4: Correspondence between α and γ in the failure-repair benchmark

C.2 Why does the simple scheme often fail?

The reason is that a long sequence of successes (or failures) tend to reduce too drastically the width of the confidence interval. Any failure (or success) posterior to that sequence would provoke a significant change of width. In practice, it would be thus necessary to keep sampling until these changes of width have a minor impact. More formally, one can indeed claim that, given a number of successes m, a coverage $1-\delta$ and a fixed width d, there exists a number of samples n such that a $(1-\delta)$-confidence interval of fixed width d, based on m and n, has a coverage greater than $1-\delta$. But there is no theoretical result that claims that the smallest integer n such that the $(1-\delta)$-confidence interval width is lower than d satisfies this coverage.

Appendix D Benchmark

D.1 Group repair benchmark

We used a benchmark small enough (125 states) to be investigated using model checker Prism [16] to corroborate our results. The system is modelled as a continuous time Markov chain and comprises three types $(1, 2, 3)$ of 4 components that may fail independently. The components fail with rates $(\lambda_1 = \alpha^2, \lambda_2 = \alpha, \lambda_3 = \alpha)$ and are repaired with rate $\mu = 1$. In addition, components are repaired with priority according to their type (type i has highest priority than type j if $i < j$). Components of type 1 and 2 are repaired simultaneously if at least two of their own type have failed. Type 3 components are repaired one by one as soon as one has failed. The property we consider is the probability γ of reaching a failure state that corresponds to the failure of all components, before returning to the initial state of no failures. Since γ is impacted by the failure rates of the system, we first determined with Prism the failure rates α such that $\gamma \in \{0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9\}$ and summarised the correspondence between α and γ in Table 4.

D.2 Prism code

We give below the code of the Prism model and the property under investigation.

```
ctmc

const int n=4;
const double alpha = 0.2384;
const double alpha2 = alpha*alpha;
const double mu = 1.0;

module type1
```

```
state1 : [0..n] init 0;
[] state1 < n -> (n-state1)*alpha2 : (state1'=state1+1);
[] state1 >=2 -> mu : (state1'=0);
endmodule

module type2
state2 : [0..n] init 0;
[] state2 < n -> (n-state2)*alpha : (state2'=state2+1);
[] state2 >=2 & state1 < 2 -> mu : (state2'=0);
endmodule

module type3
state3 : [0..n] init 0;
[] state3 < n -> (n-state3)*alpha : (state3'=state3+1);
[] state3 > 0 & state2 < 2 & state1 < 2 -> mu : (state3'=state3-1);
endmodule

label "failure" = state1 = n & state2 = n & state3 = n;
```

The property code is:

```
P=?["init" & (X !"init" U "failure")]
```

D.3 Other benchmarks

The following benchmarks are described in [24] and the Prism codes are available online[4]. For each example, we indicate below the specific parameters of the models and the property of interest.

Tandem Queueing Network In this benchmark, we set the queue capacity $c = 3$. γ is the probability that the network becomes full in T time units with $T = 15$. This probability, equal to 0.155132, is expressed in the Prism model by:

```
P=? [ true U<=T sc=c&sm=c&ph=2 ]
```

Fulfilling Specification 3 with $(\epsilon, \delta) = (0.01, 0.001)$ requires 38005 samples according to the Okamoto bound (APMC method in Prism) and about 22355 samples in average according to our method.

Fulfilling Specification 4 with $(\epsilon, \delta) = (0.05, 0.001)$ requires 194553 samples according to Dagum's algorithm and about 37558 samples in average according to our method.

Symmetric Polling System In this benchmark, we set the number of stations $N = 10$. γ is the probability, from the inital state, that station 1 is served before station 2. This probability, equal to 0.540786, is expressed in the Prism model by:

[4] http://www.prismmodelchecker.org/casestudies/index.php

```
P=? [ !(s=2&a=1) U (s=1&a=1) ]
```

Fulfilling Specification 3 with $(\epsilon, \delta) = (0.001, 0.01)$ requires 2649159 samples according to the Okamoto bound and our method (because probability γ is close to $1/2$).

Fulfilling Specification 4 with $(\epsilon, \delta) = (0.01, 0.01)$ requires 379585 samples according to Dagum's algorithm and about 103995 samples in average according to our method.

Dependable Workstation Cluster In this benchmark, we set the number of stations $N = 20$. γ is the probability, from the initial state, that QoS drops below minimum quality within T time units with $T = 1000$. This probability, equal to 5.160834×10^{-4}, is expressed in the Prism model by:

```
P=? [ true U<=T !"minimum" ]
```

Fulfilling Specification 3 with $(\epsilon, \delta) = (10^{-4}, 0.05)$ requires 184443973 samples according to the Okamoto bound and about 462265 samples in average according to our method.

Fulfilling Specification 4 with $(\epsilon, \delta) = (0.2, 0.05)$ requires 4246711 samples according to Dagum's algorithm and about 471856 samples in average according to our method.

Appendix E More results

E.1 Absolute Error schemes

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	1	0.99	0.99	0.955	0.92	0.94
$\hat{\gamma}$ min	0	0	0	0.029	0.074	0.267	0.47
$\hat{\gamma}$ max	0.019	0.024	0.039	0.072	0.122	0.326	0.533
$\hat{\gamma}$ mean	0.005	0.009	0.019	0.048	0.1	0.3	0.5
N mean	228	266	360	615	1024	2121	2450
N min	183	183	183	451	827	1994	2448
N max	364	411	539	810	1184	2208	2451
$\sigma(N)$	41	55	63	72	70	41	1

Table 5: Results of the naive scheme with $\epsilon = 0.02$ and $\delta = 0.05$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	1	1	0.985	0.99	0.99	0.985
$\hat{\gamma}$ min	0	0	0.003	0.027	0.079	0.281	0.479
$\hat{\gamma}$ max	0.013	0.024	0.035	0.067	0.121	0.322	0.525
$\hat{\gamma}$ mean	0.004	0.009	0.02	0.05	0.1	0.3	0.5
N mean	333	414	591	1041	1729	3629	4195
N min	263	263	318	703	1459	3501	4193
N max	476	660	826	1292	1994	3753	4197
$\sigma(N)$	60	82	104	104	91	49	1

Table 6: Results of the naive scheme with $\epsilon = 0.02$ and $\delta = 0.01$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	0.965	0.94	0.96	0.965	0.975	0.945
$\hat{\gamma}$ min	0	0	0.007	0.036	0.087	0.288	0.484
$\hat{\gamma}$ max	0.013	0.021	0.029	0.062	0.113	0.316	0.513
$\hat{\gamma}$ mean	0.004	0.009	0.02	0.05	0.1	0.3	0.5
N mean	518	729	1107	2172	3777	8278	9703
N min	368	368	661	1731	3384	8098	9701
N max	859	1172	1434	2583	4158	8510	9704
$\sigma(N)$	103	147	163	166	144	71	1

Table 7: Results of the naive scheme with $\epsilon = 0.01$ and $\delta = 0.05$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	0.99	0.985	0.98	1	0.99	0.985
$\hat{\gamma}$ min	0	0	0.009	0.035	0.091	0.289	0.491
$\hat{\gamma}$ max	0.013	0.016	0.029	0.06	0.109	0.309	0.513
$\hat{\gamma}$ mean	0.004	0.009	0.019	0.05	0.099	0.3	0.5
N mean	833	1157	1795	3670	6418	14226	16686
N min	528	528	1167	2761	5960	13837	16684
N max	1422	1604	2395	4282	6898	14446	16687
$\sigma(N)$	167	196	225	224	191	102	1

Table 8: Results of the naive scheme with $\epsilon = 0.01$ and $\delta = 0.01$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	1	0.995	1	0.995	1	0.995
$\hat{\gamma}$ min	0	0	0	0.031	0.083	0.282	0.487
$\hat{\gamma}$ max	0.013	0.022	0.032	0.065	0.121	0.317	0.521
$\hat{\gamma}$ mean	0.004	0.009	0.019	0.05	0.1	0.3	0.5
N mean	320	402	575	1090	1858	3972	4610
N min	243	243	243	779	1599	3842	4608
N max	470	632	804	1335	2133	4079	4612
$\sigma(N)$	62	86	108	114	103	48	1

Table 9: Results of Chen's scheme with $\epsilon = 0.02$ and $\delta = 0.05$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	1	1	1	1	1	1
$\hat{\gamma}$ min	0	0	0.004	0.034	0.083	0.281	0.484
$\hat{\gamma}$ max	0.013	0.02	0.028	0.062	0.113	0.313	0.517
$\hat{\gamma}$ mean	0.004	0.009	0.019	0.049	0.1	0.299	0.499
N mean	463	584	828	1560	2664	5691	6621
N min	349	349	461	1188	2305	5500	6619
N max	684	853	1045	1836	2931	5829	6623
$\sigma(N)$	84	102	123	124	117	63	1

Table 10: Results of Chen's scheme with $\epsilon = 0.02$ and $\delta = 0.01$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	0.98	1	0.995	1	0.995	0.995
$\hat{\gamma}$ min	0	0	0.011	0.04	0.091	0.292	0.492
$\hat{\gamma}$ max	0.01	0.017	0.028	0.059	0.107	0.31	0.511
$\hat{\gamma}$ mean	0.004	0.009	0.02	0.05	0.1	0.3	0.5
N mean	810	1171	1900	3946	7035	15684	18444
N min	489	489	1278	3257	6509	15448	18442
N max	1206	1706	2436	4503	7436	15976	18445
$\sigma(N)$	170	229	237	228	183	99	1

Table 11: Results of Chen's scheme with $\epsilon = 0.01$ and $\delta = 0.05$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	1	1	1	0.995	0.995	0.995
$\hat{\gamma}$ min	0	0.004	0.011	0.043	0.092	0.289	0.489
$\hat{\gamma}$ max	0.011	0.016	0.027	0.057	0.111	0.309	0.509
$\hat{\gamma}$ mean	0.005	0.009	0.019	0.05	0.101	0.3	0.5
N mean	1213	1685	2711	5649	10142	22540	26490
N min	702	1154	1796	5043	9414	22053	26487
N max	1796	2359	3447	6285	10977	22900	26492
$\sigma(N)$	201	229	290	248	256	125	1

Table 12: Results of Chen's scheme with $\epsilon = 0.01$ and $\delta = 0.01$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	1	1	0.99	0.995	1	0.99
$\hat{\gamma}$ min	0	0	0.003	0.029	0.082	0.281	0.476
$\hat{\gamma}$ max	0.016	0.021	0.033	0.065	0.122	0.32	0.514
$\hat{\gamma}$ mean	0.004	0.009	0.019	0.049	0.1	0.3	0.5
N mean	321	411	620	1201	2131	4609	4612
N min	243	243	304	818	1819	4504	4612
N max	556	658	885	1511	2509	4612	4612
$\sigma(N)$	70	98	110	133	110	14	0

Table 13: Results of our scheme with $\epsilon = 0.02$ and $\delta = 0.05$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	1	1	1	1	1	1
$\hat{\gamma}$ min	0	0	0.002	0.036	0.084	0.286	0.485
$\hat{\gamma}$ max	0.016	0.021	0.032	0.063	0.114	0.313	0.52
$\hat{\gamma}$ mean	0.004	0.01	0.019	0.049	0.1	0.3	0.5
N mean	456	596	869	1673	2930	6405	6623
N min	332	332	402	1325	2548	6245	6623
N max	765	909	1215	2036	3244	6543	6623
$\sigma(N)$	88	113	137	140	134	65	0

Table 14: Results of our scheme with $\epsilon = 0.02$ and $\delta = 0.01$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	0.99	0.995	0.995	0.995	1	1
$\hat{\gamma}$ min	0	0	0.01	0.039	0.089	0.291	0.491
$\hat{\gamma}$ max	0.011	0.019	0.027	0.059	0.106	0.309	0.51
$\hat{\gamma}$ mean	0.004	0.009	0.019	0.05	0.1	0.3	0.5
N mean	831	1229	2064	4474	8161	18434	18445
N min	448	448	1251	3571	7427	18221	188445
N max	1401	2014	2690	5180	8612	18445	18445
$\sigma(N)$	194	263	250	278	223	115	0

Table 15: Results of our scheme with $\epsilon = 0.01$ and $\delta = 0.05$

γ	0.005	0.01	0.02	0.05	0.1	0.3	0.5
Coverage	1	1	1	1	0.995	1	1
$\hat{\gamma}$ min	0	0.001	0.013	0.044	0.089	0.292	0.492
$\hat{\gamma}$ max	0.01	0.015	0.026	0.057	0.109	0.309	0.508
$\hat{\gamma}$ mean	0.005	0.009	0.02	0.05	0.1	0.3	0.5
N mean	1192	1738	2908	6198	11266	25380	26492
N min	667	808	2112	5552	10181	24990	26492
N max	1759	2351	3573	6961	12057	25791	26492
$\sigma(N)$	228	260	304	293	289	140	0

Table 16: Results of our scheme with $\epsilon = 0.01$ and $\delta = 0.01$

E.2 Relative Error schemes

γ	0.9	0.7	0.5	0.3	0.1	0.05	0.01	0.001
Coverage (New)	0.98	0.99	1	1	1	0.99	1	0.99
N mean Dagum	1360	3160	6412	14253	52432	111703	570763	5787456
N mean Watanabe	1353	1738	2439	4048	12207	24362	122029	1218779
N mean New	137	441	991	2204	8208	17356	88838	895496
N min Dagum	1239	2706	4893	10515	38047	82396	402704	3842171
N min Watanabe	1326	1667	2290	3784	11335	22356	112202	1124053
N min New	59	321	773	1931	7387	15992	79697	826619
N max Dagum	1537	3832	8213	19296	71621	153484	919717	8408400
N max Watanabe	1381	1808	2579	4371	12995	27058	131959	1303185
N max New	230	558	1163	2425	8917	18836	97401	976689
$\sigma(N)$ Dagum	51	220	641	1591	6414	14638	75480	803964
$\sigma(N)$ Watanabe	11	27	52	102	351	751	3436	31828
$\sigma(N)$ New	32	41	64	92	301	549	2921	28891

Table 17: Results of the Relative Error schemes with $\epsilon = 0.1$ and $\delta = 0.05$

γ	0.9	0.7	0.5	0.3	0.1	0.05	0.01	0.001
N mean Dagum	1871	4402	9056	19703	74064	152757	803572	8124356
N mean Watanabe	1942	2498	3501	5836	17479	35006	175092	1746713
N mean New	202	623	1373	3043	11365	23812	122426	1236491
N min Dagum	1749	3784	7344	14442	47525	110121	544802	6765298
N min Watanabe	1900	2404	3356	5525	16320	31990	165058	1630908
N min New	101	474	1188	2777	10549	21975	112569	1144960
N max Dagum	2068	5042	11530	24048	99876	199447	1088138	9218545
N max Watanabe	1989	2602	3682	6244	18857	37443	186327	1871956
N max New	288	752	1593	3398	12513	25644	132023	1334497
$\sigma(N)$ Dagum	57	267	766	1957	9002	17235	101350	757926
$\sigma(N)$ Watanabe	16	36	57	125	420	865	4023	47309
$\sigma(N)$ New	37	54	66	114	338	749	3479	36934

Table 18: Results of the Relative Error schemes with $\epsilon = 0.1$ and $\delta = 0.01$

γ	0.9	0.7	0.5	0.3	0.1	0.05	0.01	0.001
N mean Dagum	3162	8912	19244	43276	163084	346269	1800585	18208080
N mean Watanabe	5163	6634	9299	15453	46496	92950	465144	4650289
N mean New	476	1631	3737	8473	32175	67850	348706	3515688
N min Dagum	2744	7655	15878	32973	118648	263215	1288876	14979100
N min Watanabe	5093	6490	9019	14991	44740	89176	445042	4418313
N min New	331	1422	3318	7988	30794	64989	341699	3472015
N max Dagum	3617	10336	23284	54222	213356	466642	2348861	21291654
N max Watanabe	5250	6813	9551	15896	48279	97027	481377	4879618
N max New	634	1879	4059	8950	33861	70469	359423	3588372
$\sigma(N)$ Dagum	156	552	1497	3882	15851	35758	196226	2262304
$\sigma(N)$ Watanabe	25	50	92	168	645	1383	6796	67340
$\sigma(N)$ New	61	91	109	195	624	1124	5953	37749

Table 19: Results of the Relative Error schemes with $\epsilon = 0.05$ and $\delta = 0.05$

γ	0.9	0.7	0.5	0.3	0.1	0.05	0.01	0.001
N mean Dagum	4394	12337	26677	60263	226889	479164	2467430	25300472
N mean Watanabe	7416	9540	13347	22235	66756	133581	665536	6677525
N mean New	669	2266	5151	11675	44346	93236	482998	4871059
N min Dagum	3857	10779	22533	45883	183403	369946	1806323	22380395
N min Watanabe	7352	9350	13024	21692	65096	129667	658412	6507035
N min New	418	1965	4815	11026	42615	89031	461921	4670509
N max Dagum	4925	14566	31594	72762	290120	601099	3148179	27905975
N max Watanabe	7482	9731	13674	22775	68610	139543	675949	6824914
N max New	876	2548	5523	12331	46301	97555	503316	5085098
$\sigma(N)$ Dagum	190	675	1795	4909	20553	45161	228655	2010892
$\sigma(N)$ Watanabe	29	68	119	214	748	1573	6733	89540
$\sigma(N)$ New	68	106	142	245	689	1418	7170	73074

Table 20: Results of the Relative Error schemes with $\epsilon = 0.05$ and $\delta = 0.01$