A variation of the Newton-Pepys problem and its connections to size-estimation problems

Damiano Varagnolo[a,*], Luca Schenato[b], Gianluigi Pillonetto[b]

[a] *School of Electrical Engineering, KTH Royal Institute of Technology, Osquldas V ag 10, SE-100 44 Stockholm, Sweden*
[b] *Department of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131 Padova, Italy*

Abstract

This paper considers a variation of the 17^{th} century problem commonly known as the Newton-Pepys problem, or the John Smith's problem. We provide its solution and interpret the result in terms of maximum likelihood estimation. In addition, we illustrate the practical relevance of these findings for solving size-estimation problems, and in particular for determining the number of agents in a wireless sensor network.

Keywords: Bernoulli trials, John Smith problem

1. Introduction

In November and December 1693, Samuel Pepys asked, with some letters to Isaac Newton, a question involving gambling (Turnbull, 1961, pp. 293–303), (Pepys, 1929, Vol. 1, pp. 72–94). The question, after a slight reformulation, was which of the following events is most likely to happen:

1. at least one "6" appears when six fair dice are tossed independently;
2. at least two "6"s appear when twelve fair dice are tossed independently;
3. at least three "6"s appears when eighteen fair dice are tossed independently.

Originally posed by a colleague of Pepys, named John Smith, and collected in the famous and posthumously published Pepys' diaries, the question is described from a mathematical point of

*Corresponding author
Email addresses: damiano@kth.se (Damiano Varagnolo), schenato@dei.unipd.it (Luca Schenato), giapi@dei.unipd.it (Gianluigi Pillonetto)

view by several papers[1], e.g., David (1957); Rubin and Evans (1961); Rubin and Schell (1960). Here we consider a variation of a generalized version of this problem and then show how it can be used to solve modern size-estimation problems.

2. The generalized Newton-Pepys problem and the solutions known from the literature

Let the experiment consist in throwing n non-necessarily fair dice, each with a given number of faces and with probabilities that are uniform among the dice. Let p denote the probability that a single die will select a certain face when thrown. The number of faces of the dice and which face should be selected are irrelevant for our results. Let r be the random variable "total number of dice that selected the considered face". Let $\text{pr}\left[r \geq k\ ;\ n, p\right]$ be the probability of having at least k correct selections when throwing exactly n dice.

The following question generalizes the original Newton-Pepys problem: let $\nu_1, \nu_2 \in \mathbb{N}_+$, $\nu_1 \leq \nu_2$.

$$\textit{Is } \text{pr}\left[r \geq \nu_1 k\ ;\ \nu_1 n, p\right] \textit{ not smaller than } \text{pr}\left[r \geq \nu_2 k\ ;\ \nu_2 n, p\right] \textit{ for all } n,\ p,\ k\textit{?} \tag{1}$$

The original Newton-Pepys problem is a specific instance of (1), since it can be translated into "is $\text{pr}\left[r \geq 1\ ;\ 6, 1/6\right] \geq \text{pr}\left[r \geq 2\ ;\ 12, 1/6\right] \geq \text{pr}\left[r \geq 3\ ;\ 18, 1/6\right]$?".

Answers to the newly posed problem depend on the particular instances of k, n and p. As noticed in Rubin and Evans (1961), in fact, between $\text{pr}\left[r \geq \nu_1 k\ ;\ \nu_1 n, p\right]$ and $\text{pr}\left[r \geq \nu_2 k\ ;\ \nu_2 n, p\right]$ there are no uniform relationships in k, n and p, and the former may be bigger or smaller than the latter depending on the particular instance. E.g.,

$$\begin{aligned} 0.75 = \text{pr}\left[r \geq 1\ ;\ 2, 0.5\right] &> \text{pr}\left[r \geq 2\ ;\ 4, 0.5\right] = 0.6875 \\ 0.875 = \text{pr}\left[r \geq 1\ ;\ 3, 0.5\right] &< \text{pr}\left[r \geq 2\ ;\ 6, 0.5\right] = 0.8906... \end{aligned} \tag{2}$$

Nonetheless, we can find results on particular scenarios from Chaundy and Bullard (1960), as listed below.

Proposition 1. *If* $k_1, k_2, n \in \mathbb{N}_+$ *and* $k_1 < k_2$ *then*

$$\text{pr}\left[r \geq k_1\ ;\ k_1 n, \frac{1}{n}\right] > \text{pr}\left[r \geq k_2\ ;\ k_2 n, \frac{1}{n}\right]\ . \tag{3}$$

[1] As an hystorical note, Newton's answer was analytically correct despite the underlying logical motivations were wrong, see Stigler (2006).

Notice that the above proposition solves the Newton-Pepys problem, showing that the first of the three scenarios mentioned in the Introduction is the most likely event. The following result also holds.

Proposition 2. *If* $k, n_1, n_2 \in \mathbb{N}_+$ *and* $n_1 < n_2$ *then*

$$\mathrm{pr}\left[r \geq k \ ; \ kn_1, \frac{1}{n_1}\right] > \mathrm{pr}\left[r \geq k \ ; \ kn_2, \frac{1}{n_2}\right] \ . \tag{4}$$

3. A variation of the Newton-Pepys problem and its solution

The following problem is a variation of that contained in Section 2: let $\nu_1, \nu_2 \in \mathbb{N}_+$, $\nu_1 \leq \nu_2$.

$$\textit{Is } \mathrm{pr}\left[r = \nu_1 k \ ; \ \nu_1 n, p\right] \textit{ not smaller than } \mathrm{pr}\left[r = \nu_2 k \ ; \ \nu_2 n, p\right] \textit{ for all } n, \ p? \tag{5}$$

This question is different from its ancestor in Section 1. The latter, in fact, under this scenario would become which of the following events is most likely to happen:

1. exactly one "6" appears when six fair dice are tossed independently;
2. exactly two "6"s appear when twelve fair dice are tossed independently;
3. exactly three "6"s appears when eighteen fair dice are tossed independently.

However, differently from (1), (5) has always a positive answer as stated below.

Proposition 3. *If* $\nu_1, \nu_2, n, k \in \mathbb{N}_+$, $\nu_1 \leq \nu_2$, $k \leq n$ *and* $p \in [0, 1]$ *then*

$$\mathrm{pr}\left[r = \nu_1 k \ ; \ \nu_1 n, \ p\right] \geq \mathrm{pr}\left[r = \nu_2 k \ ; \ \nu_2 n, \ p\right] \ . \tag{6}$$

Proof. Equivalences

$$\mathrm{pr}\left[r = \nu_1 k \ ; \ \nu_1 n, \ p\right] = \binom{\nu_1 n}{\nu_1 k} p^{\nu_1 k} (1-p)^{\nu_1 (n-k)}$$

$$\mathrm{pr}\left[r = \nu_2 k \ ; \ \nu_2 n, \ p\right] = \binom{\nu_2 n}{\nu_2 k} p^{\nu_2 k} (1-p)^{\nu_2 (n-k)}$$

imply that (6) can be rewritten as

$$\frac{\binom{\nu_2 n}{\nu_2 k}}{\binom{\nu_1 n}{\nu_1 k}} p^{(\nu_2-\nu_1)k}(1-p)^{(\nu_2-\nu_1)(n-k)} \leq 1 \; .$$

It is straightforward to check that $p^{(\nu_2-\nu_1)k}(1-p)^{(\nu_2-\nu_1)(n-k)}$ is concave for $p \in [0,1]$. Thus from

$$\begin{aligned} \frac{\partial}{\partial p}\left(p^{(\nu_2-\nu_1)k}(1-p)^{(\nu_2-\nu_1)(n-k)}\right) = \\ = (\nu_2-\nu_1)p^{(\nu_2-\nu_1)k-1}(1-p)^{(\nu_2-\nu_1)(n-k)-1}(k-np) \end{aligned} \tag{7}$$

we obtain

$$p^{(\nu_2-\nu_1)k}(1-p)^{(\nu_2-\nu_1)(n-k)} \leq \left(\frac{k}{n}\right)^{(\nu_2-\nu_1)k} \left(\frac{n-k}{n}\right)^{(\nu_2-\nu_1)(n-k)}$$

and this implies that the condition for (6) to hold is

$$\frac{\binom{\nu_2 n}{\nu_2 k}}{\binom{\nu_1 n}{\nu_1 k}} \left(\frac{k}{n}\right)^{(\nu_2-\nu_1)k} \left(\frac{n-k}{n}\right)^{(\nu_2-\nu_1)(n-k)} \leq 1 \; . \tag{8}$$

Considering that

$$\frac{\binom{\nu_2 n}{\nu_2 k}}{\binom{\nu_1 n}{\nu_1 k}} = \frac{(\nu_2 n)!}{(\nu_1 n)!}\frac{(\nu_1 k)!}{(\nu_2 k)!}\frac{(\nu_1(n-k))!}{(\nu_2(n-k))!} = \frac{a}{b\cdot c} = \frac{d}{e\cdot f}$$

with

$$b := (\nu_2 k)\cdot(\nu_2 k-1)\cdot \ \ldots \ \cdot(\nu_1 k+1) \tag{9}$$

$$c := \big(\nu_2(n-k)\big)\cdot\big(\nu_2(n-k)-1\big)\cdot \ \ldots \ \cdot(\nu_1(n-k)+1) \tag{10}$$

$$a := (\nu_2 n)\cdot(\nu_2 n-1)\cdot \ \ldots \ \cdot(\nu_1 n+1) \tag{11}$$

and

$$\begin{aligned} e &:= (k)\cdot\left(k-\frac{1}{\nu_2}\right)\cdot \ \ldots \ \cdot\left(k-\frac{(\nu_2-\nu_1)k-1}{\nu_2}\right) \\ f &:= (n-k)\cdot\left(n-k-\frac{1}{\nu_2}\right)\cdot \ \ldots \ \cdot\left(n-k-\frac{(\nu_2-\nu_1)(n-k)-1}{\nu_2}\right) \\ d &:= (n)\cdot\left(n-\frac{1}{\nu_2}\right)\cdot \ \ldots \ \cdot\left(n-\frac{(\nu_2-\nu_1)n-1}{\nu_2}\right) \; . \end{aligned}$$

We notice that b, e are the product of $(\nu_2-\nu_1)k$ terms, c, f are the product of $(\nu_2-\nu_1)(n-k)$ terms, and a, d are the product of $(\nu_2-\nu_1)n$ terms. We can thus rewrite condition (8) as

$$\frac{k^{(\nu_2-\nu_1)k}}{e}\cdot\frac{(n-k)^{(\nu_2-\nu_1)(n-k)}}{f}\leq\frac{n^{(\nu_2-\nu_1)n}}{d}$$

or, equivalently, as

$$\left[\prod_{i=1}^{(\nu_2-\nu_1)k}\frac{k}{k-\frac{i-1}{\nu_2}}\right]\cdot\left[\prod_{j=1}^{(\nu_2-\nu_1)(n-k)}\frac{(n-k)}{(n-k)-\frac{j-1}{\nu_2}}\right]\leq\left[\prod_{\ell=1}^{(\nu_2-\nu_1)n}\frac{n}{n-\frac{\ell-1}{\nu_2}}\right]\ . \tag{12}$$

Remarkably, the inequality before can be proved obtaining a much stronger result involving terms-by-terms inequalities. More precisely, a sufficient condition for (12) to hold is that, assuming that both sides are sorted, each ℓ-th term in the right-hand-side is not smaller than ℓ terms in the left-hand-side. Then, considering the inequalities

$$\frac{k}{k-\frac{i-1}{\nu_2}}\ \leq\ \frac{n}{n-\frac{\ell-1}{\nu_2}}\qquad i\in\{1,\dots,(\nu_2-\nu_1)k\} \tag{13}$$

$$\frac{(n-k)}{(n-k)-\frac{j-1}{\nu_2}}\ \leq\ \frac{n}{n-\frac{\ell-1}{\nu_2}}\qquad j\in\{1,\dots,(\nu_2-\nu_1)(n-k)\}\ , \tag{14}$$

a sufficient condition for (12) is that for every $\ell\in\{1,\dots,(\nu_2-\nu_1)n\}$ the number of i's satisfying (13) plus the number of j's satisfying (14) has to be at least ℓ. It is convenient to rewrite (13) and (14) as

$$i\leq\frac{k}{n}(\ell-1)+1\qquad i\in\{1,\dots,(\nu_2-\nu_1)k\} \tag{15}$$

$$j\leq\frac{(n-k)}{n}(\ell-1)+1\qquad j\in\{1,\dots,(\nu_2-\nu_1)(n-k)\}\ . \tag{16}$$

If $\lfloor\cdot\rfloor$ denotes the floor operator, the sum of the number of i's satisfying (15) and the number of j's satisfying (16) is thus given by

$$\left\lfloor\frac{k}{n}(\ell-1)+1\right\rfloor+\left\lfloor\frac{(n-k)}{n}(\ell-1)+1\right\rfloor\ .$$

We eventually conclude that (6) is satisfied if

$$\left\lfloor\frac{k}{n}(\ell-1)+1\right\rfloor+\left\lfloor\frac{(n-k)}{n}(\ell-1)+1\right\rfloor\geq\ell\qquad\forall\ell\in\{1,\dots,(\nu_2-\nu_1)n\}\ . \tag{17}$$

To prove (17), we notice that, since if $x,y\in\mathbb{R}$ then $\lfloor x\rfloor+\lfloor y\rfloor\geq\lfloor x+y\rfloor-1$,

$$\begin{aligned}\left\lfloor\frac{k}{n}(\ell-1)+1\right\rfloor+\left\lfloor\frac{(n-k)}{n}(\ell-1)+1\right\rfloor\geq\\ \geq\left\lfloor\frac{k}{n}(\ell-1)+1+\frac{(n-k)}{n}(\ell-1)+1\right\rfloor-1\ .\end{aligned}$$

Exploiting now the fact that if $x \in \mathbb{R}, y \in \mathbb{N}$ then $\lfloor x \rfloor + y = \lfloor x \rfloor + \lfloor y \rfloor = \lfloor x + y \rfloor$, we can conclude that (17) holds true for all $\ell \in \mathbb{N}_+$ since

$$\left\lfloor \frac{k}{n}(\ell-1)+1 \right\rfloor + \left\lfloor \frac{(n-k)}{n}(\ell-1)+1 \right\rfloor \geq$$
$$\geq \left\lfloor \frac{k}{n}(\ell-1)+1+\frac{(n-k)}{n}(\ell-1) \right\rfloor = \lfloor \ell \rfloor = \ell \, .$$

This thus proves the result. □

It is worth noticing that the proof of Proposition 3 is not based on induction concepts and is essentially different from the proofs of Propositions 1 and 2.

4. Practical implications related to size-estimation problems

Assume that a certain experiment, run an unknown number of times, had a success fraction $f = k/n$ with k, number of successes, and n, number of runs, both unknown. Assume that $k/n = \widehat{k}/\widehat{n}$, with $\widehat{k}$ and $\widehat{n}$ coprime. Proposition 3 implies that $\text{pr}\left[r = \widehat{k}\,;\, \widehat{n},\, p\right] \geq \text{pr}\left[r = 2\widehat{k}\,;\, 2\widehat{n},\, p\right] \geq \ldots$, for all possible p's. In other words, the numerator and denominator of the coprime representation of f are the Maximum Likelihood (ML) estimates of the number of successes and of trials, respectively, irrespective of the actual probability p of success of the single trial. This is due to the fact that, the larger n is, the more the binomial distributions are spread out, so that the individual probabilities are smaller. An alternative interpretation is in terms of Ockham's razor: the simplest hypothesis, i.e., the one invoking the fewest trials, is also the most likely one.

Remarkably, this fact, answering a variation of a so ancient question, has interesting implications in modern problems. In particular, let us now focus on distributed estimation of the size of a wireless sensor network Akyildiz et al. (2002): the aim is to determine the number of collaborating agents, which is a useful information for maintenance and organization purposes. Now, let n denote the unknown number of agents. We propose a strategy to estimate n that can be used in anonymous networks, i.e., networks where agents are not assured to have unique IDs or are not allowed to disclose them due to privacy reasons. This problem is intricate in view of the following impossibility result:

Theorem 4 (Theorem 9 in Cidon and Shavitt (1995))**.** *There exists no algorithm that is able to compute the size of a generic network of anonymous agents that terminates with the correct result*

for every finite execution with probability one, and that has a bounded average bit complexity (i.e., s.t. the average number of bits used by the algorithm is bounded).

The aim is then to design probabilistic algorithms with the smallest (but unavoidably non-null) probability of error.

4.1. Size-estimation under infinite-precision arithmetics

Proposition 3 suggests the following strategy: let each agent $i = 1, \dots, n$ locally draw a single $y_i \in \{0, 1\}$ from independent Bernoulli random variables of parameter p. Under the stated assumptions, $f = \sum_{i=1}^{n} y_i/n = k/n$ is the fraction of ones generated by the various agents.

Let then the agents distributedly compute the exact $f = k/n$ as follows[2]: every $i = 1, \dots, n$ has a local variable $f_i(\tau)$ (τ denotes time) initialized as $f_i(0) = y_i$. Then every second every $i = 1, \dots, n$ randomly selects one of its neighbors (say, j). Thus both i and j average their local variables, i.e., let $f_i(\tau+1) = \frac{1}{2}\big(f_i(\tau)+f_j(\tau)\big)$, $f_j(\tau+1) = \frac{1}{2}\big(f_i(\tau)+f_j(\tau)\big)$. With this scheme $\lim_{\tau\to+\infty} f_i(\tau) = f$ with probability 1, exponentially in time and for every $i = 1, \dots, n$ (Fagnani and Zampieri, 2008, Example 3.4).

Let us assume for now that each agent has eventually computed the *asymptotic exact* $f = k/n = \widehat{k}/\widehat{n}$ with $\widehat{k}$ and $\widehat{n}$ coprime (relaxations of this hypothesis will be analyzed consequently). As noticed at the beginning of section 4, $\widehat{n}$ is the ML estimate of n. Remarkably, n is a multiple of $\widehat{n}$ by construction , implying $\widehat{n} \leq n$. Moreover $\widehat{n}$ is correct, i.e., $\widehat{n} = n$, if and only if k is a totative of n, i.e., if k is coprime with n. The number of k's leading to correct estimates is thus equal to the number of totatives of n, known as the Euler's ϕ-function $\phi(n)$ Lehmer (1955).

To increase the statistical performance of the estimator we can let the agents perform M experiments in parallel, i.e., extract $y_{i,m} \in \{0, 1\}$, $i = 1, \dots, n$, $m = 1, \dots, M$ still from i.i.d. Bernoulli random variables, then compute $f_m = \frac{1}{n}\sum_{i=1}^{n} y_{i,m}$ as above, and eventually compute $\widehat{n}_m$ from f_m as above. Let then $\mathrm{LCM}(\cdot)$ indicate the least common multiple operator. Since n is a common multiple of all the $\widehat{n}_m$, $n \geq \mathrm{LCM}(\widehat{n}_1, \dots, \widehat{n}_M)$. Thanks to the monotonicity of the likelihoods presented

[2]This scheme is the symmetric gossip version of the *average consensus* Xiao et al. (2007) Garin and Schenato (2011), a famous distributed algorithm tailored for the computation of averages. We refer to it just for sake of clarity and because of its little coordination requirements and simplicity. The following results indeed do not depend on how $f = k/n$ is actually computed.

in Proposition 3, the following result holds.

Proposition 5. *The ML estimator for n given $f_1, \dots, f_M$ is*

$$\widehat{n} = LCM(\widehat{n}_1, \dots, \widehat{n}_M)\,. \tag{18}$$

Moreover,

$$\mathrm{pr}\,[n \neq \widehat{n}] \leq \prod_{m=1}^{M} \mathrm{pr}\,[n \neq \widehat{n}_m] = \left(\mathrm{pr}\,[n \neq \widehat{n}_1]\right)^M \tag{19}$$

thus the probability of error decays exponentially in M, uniformly on the Bernoulli parameter p.

Proof. • *ML property:* assume the knowledge of just a single f_m. Since

$$\mathrm{pr}\,[f_m\ ;\ n] = \binom{n}{nf_m} p^{nf_m}(1-p)^{n-nf_m},$$

the likelihood $\mathrm{pr}\,[f_m\ ;\ n]$ is non-null only for $n \in \mathcal{I}_m := \{\widehat{n}_m, 2\widehat{n}_m, 3\widehat{n}_m, \dots\}$. Moreover, since the $y_{i,m}$'s are independent,

$$\mathrm{pr}\,[f_1, \dots, f_M\ ;\ \overline{n}] = \prod_{m=1}^{M} \mathrm{pr}\,[f_m\ ;\ \overline{n}]\,.$$

This eventually means $\mathrm{pr}\,[f_1, \dots, f_M\ ;\ n]$ to be non-null only for $n \in \mathcal{I} := \cap_{m=1}^{M} \mathcal{I}_m$.

Now, Proposition 3 implies $\mathrm{pr}\,[f_m\ ;\ n]$ restricted to $\mathcal{I}_m$ to be non increasing, that implies $\mathrm{pr}\,[f_1, \dots, f_M\ ;\ \overline{n}]$ restricted to $\mathcal{I}$ to be non increasing. This eventually implies that the maximum likelihood estimator of n is $\widehat{n} = \mathrm{LCM}\,(\widehat{n}_1, \dots, \widehat{n}_M) = \min(\mathcal{I})$.

• *Equation* (19)*:* consider that the single $\widehat{n}_m$ is obtained reducing $f_m = k_m/n$ to a coprime fraction $\widehat{k}/\widehat{n}$. This implies that $\widehat{n}_m$ is a factor of n and thus, by construction, that each $\widehat{n}_m$ cannot overestimate n. Consider then that $\widehat{n}$ is obtained as the least common multiple of factors of n. Since the LCM of factors of a number cannot be bigger than the number itself, again by construction $\widehat{n}$ cannot overestimate n.

Consider then that if $n \neq \widehat{n}$ then necessarily $n \neq \widehat{n}_1, \dots, n \neq \widehat{n}_M$. (The contrary is not true, in the sense that but that $n \neq \widehat{n}_1, \dots, n \neq \widehat{n}_M$ does not imply $n \neq \widehat{n}$. E.g., $M = 2, n = 6, \widehat{n}_1 = 3, \widehat{n}_2 = 2$.) This thus directly implies that $\mathrm{pr}\,[n \neq \widehat{n}] \leq \mathrm{pr}\,[n \neq \widehat{n}_1, \dots, n \neq \widehat{n}_M]$, that leads to (19) since the various $\widehat{n}_m$'s are i.i.d. □

Since the error probability $\mathrm{pr}\,[n \neq \widehat{n}_1]$ is a function of p, one might desire to find the optimal value p. A possible direction is to follow the classical Fischerian approach of selecting the p that

minimizes the worst probability of error $\mathrm{pr}\,[n \neq \widehat{n}]$ over a suitable set of possible n's. It is nonetheless easy to verify through simple numerical experiments that for $M = 1$ the choice is not particularly critical and that, for any $p \in [0.25, 0.75]$ and for any $n \leq 1000$ then $\mathrm{pr}\,[n \neq \widehat{n}_1] \leq 0.85$. We therefore choose $p = 0.5$ in the simulations in next session. According to Proposition 5 where $p = 0.5$ and $n \leq 1000$, then $\mathrm{pr}\,[n \neq \widehat{n}] \leq (0.85)^M$. Hence the strategy leads to an estimator that, increasing the number of independent trials M, achieves any desirable level of confidence, with the probability of error decaying to zero exponentially fast with M.

4.2. Implementation under finite-precision arithmetics

In real devices the $f_i(\tau)$'s must be represented using a finite number of bits b. In other words, $f_i(\tau) \in \mathcal{F} \subset [0, 1]$ with $\mathcal{F}$ a finite set of 2^b points in $[0, 1]$. Importantly, some fractions k/n might not be in $\mathcal{F}$. E.g., Figure 1 represents a particular $\mathcal{F}$ for $b = 6$. In this case $1/4 \notin \mathcal{F}$.

To implement the strategy proposed in Section 4.1 in this more realistic scenario we then notice the following facts:

- assuming the knowledge of an upper bound on n, say $n_{\max}$, the set of potential fractions $\mathcal{KN} := \{f = k/n \mid n = 1, \ldots, n_{\max}, k = 0, \ldots, n\}$ is finite. This implies that the various elements of $\mathcal{F}$ can be mapped onto the closest potential fraction $k/n \in \mathcal{KN}$, and subsequently the fraction k/n onto its denominator. This operation is represented in Figure 1 by means of gray rectangles;

- if the elements of $\mathcal{F}$ are equally spaced then the previous operations $f_i(\tau+1) = \frac{1}{2}\big(f_i(\tau)+f_j(\tau)\big)$ in $\mathcal{F}$ can be refined so that every $f_i(\tau)$'s converges *in finite time* to $\widetilde{f} = f + e \neq f$, with e an error smaller than the spacing of the elements in $\mathcal{F}$ Carli et al. (2010).

Since the smallest distance between the elements in $\mathcal{KN}$ is $\dfrac{1}{n_{\max}(n_{\max}-1)}$,we implicitly obtain that if $\dfrac{1}{2^b} < \dfrac{1}{2n_{\max}(n_{\max}-1)}$ then $\widetilde{f}$ and f are mapped to the *same* fraction k/n. I.e., if the number of bits is sufficiently high then the computation of $\widehat{n}$ is insensitive to the presented finite-precision arithmetics issues, that means that (19) still holds.

4.3. Comparisons with random-walks based estimators

The performance of the proposed estimator are intrinsically different from the ones of typical probabilistic counting and anonymous size-estimation techniques, e.g., Flajolet and Martin (1985);

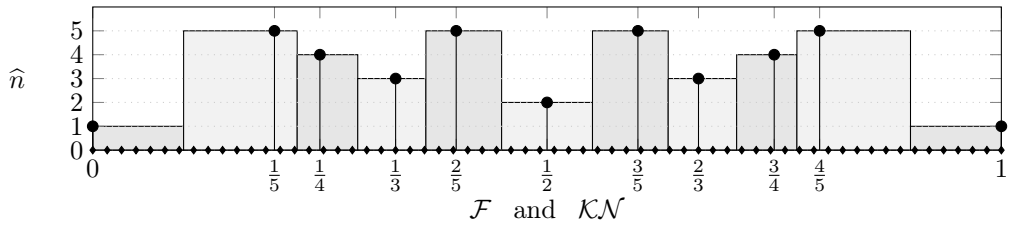


Figure 1: Set of the representable values $\mathcal{F}$ (64 diamonds over the x-axis, corresponding to $b = 6$) and of the plausible fractions $\mathcal{KN}$ (stems), among with their corresponding estimates $\widehat{n}$ for $n_{\max} = 5$. The gray rectangles indicate how each element of $\mathcal{F}$ is associated to a certain estimate $\widehat{n}$.

Massouliè et al. (2006); Sirken and Shimizu (1999). The latters, in fact, fuse independent trials by *averaging* the single outcomes, and this leads to an error variance that decays with the inverse of the number of trials, i.e. as $1/M$. Differently, the estimator proposed in Proposition 5 performs LCM operations on the single outcomes, and this leads to error probabilities decaying exponentially with the number of trials M, i.e., as $\alpha^M, \alpha < 1$.

In this section we specifically compare our strategy with one of the most used strategies, the class of the so called random-walks based estimators (see, e.g., Massouliè et al. (2006)). The inference mechanism works as follows: a querying node (e.g., the black one in Figure 2(a)) initiates the procedure generating M "batons". Then for each baton the querier randomly selects (with replacement) one of his neighbors, then add a mark to the baton and pass it to the selected neighbor – as in a relay race. The receiver, in its turn, perform the same operation: randomly select one of his neighbors, add an other mark to the baton and then pass it to the selected neighbor, and so on. Every baton thus randomly travels through the network, until it returns to the querier. The latter can then infer the network size by combining how many times each baton has been passed before returning to the original position. With this strategy the error variance decays as $1/M$.

Figure 2(b) shows a typical realization of the temporal evolution of the estimates given by the random walk strategy applied to the communication network shown in the first panel ($M = 10$). Here the querying node is the black one, and the x-axis measures how many times the batons have been propagated from a node to another one. Figure 2(c) instead shows a typical realization of the estimates obtained by the black node when applying our novel strategy to the same communication network ($M = 10$, $p = 0.5$, $n_{\max} = 50$, each scalar represented with 16 bits). The x-axis unit is the number of communications performed by the black node. Importantly, *before reaching their*

final values the various $f_{i,m}(\tau)$'s may visit several different elements of $\mathcal{F}$. They thus might be temporarily associated to fractions in $\mathcal{KN}$ whose denominators are not factors of n. E.g., consider a network of $n = 20$ agents where, at time $\tau = 100$, $f_{i,m}(100) = 1/8$. This implies that at $\tau = 100$ agent i sets $\widehat{n}_m = 8$ – not a factor of $n = 20$. Similarly, even for $M > 1$, it may thus happen that $\widehat{n} = \mathrm{LCM}\,(\widehat{n}_1, \ldots, \widehat{n}_M) > n_{\max}$. However, if the number of bits is sufficiently high, then this effect is just temporary since eventually $f_{i,m}(\tau)$ will converge to the correct f_m and thus be associated to a correct factor of n. The temporary case $\widehat{n} > n_{\max}$ has then been managed in our simulations by arbitrarily setting the unreliable estimates to zero.

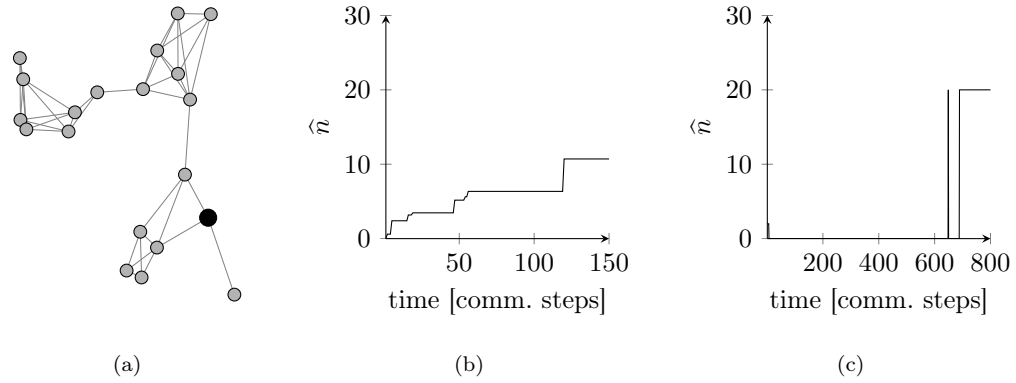
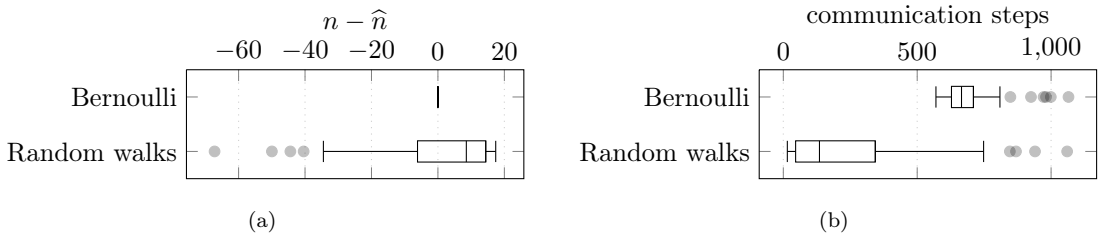


Figure 2: 2(a): communication network ($n = 20$, black node = querying node). 2(b): trajectory of estimates for random walks, $M = 10$. 2(c): trajectory of estimates for Proposition 5, $M = 10$, $p = 0.5$.

We notice the following differences between the two schemes. When using random walks, the estimates are computed by just one agent, are updated when the various seeds return to the querier and are monotonically increasing in time. When using our scheme, estimates are computed in parallel every time an agent communicates and there is no monotonicity. More importantly, errors variances of the two estimators scale differently with M: for random walks strategies it decays polynomially with M while for our strategy, thanks to Proposition 5, it decays exponentially.

To further illustrate these effects, we have performed 100 independent experiments, each running the two previously considered strategies using again the communication network of Figure 2(a), $M = 10$ and $p = 0.5$. The empirical spread of the estimation errors is plotted in Figure 3(a), while the spread of the convergence times is shown in Figure 3(b). The stopping criterion in our strategy is: stop if the current estimate is valid and has not changed in the last 10 steps. Remarkably, our

algorithm allows the agents to estimate perfectly the network size in all the 100 experiments at the price of a longer convergence time. This also indicates that the upper bound (19) (in this case, $0.85^{10} \approx 0.2$) can be very conservative.



(a) (b)

Figure 3: 3(a): estimation errors. 3(b): convergence times.

5. Conclusions

We have illustrated and solved a variation of the historical Newton-Pepys problem. In addition, we have shown the connection between this result and ML size-estimation problems, deriving a distributed strategy to estimate the number of agents composing an anonymous network. Despite the fact that the derived algorithm has desirable theoretical and practical qualities like accuracy and simplicity, its actual implementation poses interesting future research directions that we plan to investigate, e.g., related to assessing its sensitivity to round-off errors.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement n°257462 HYCON2 Network of excellence, by Progetto di Ateneo CPDA090135/09 funded by the University of Padova, by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. We would moreover thank the anonymous Reviewers and the Editor for the important suggestions and for having considerably improved the readability of the manuscript.

References

Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E., 2002. A survey on sensor networks. IEEE Communications Magazine 40, 102–114.

Carli, R., Fagnani, F., Frasca, P., Zampieri, S., 2010. Gossip consensus algorithms via quantized communication. Automatica 46, 70–80.

Chaundy, T.W., Bullard, J.E., 1960. John Smith's Problem. The Mathematical Gazette 44, 253–260.

Cidon, I., Shavitt, Y., 1995. Message terminating algorithms for anonymous rings of unknown size. Information Processing Letters 54, 111–119.

David, F., 1957. Mr Newton, Mr Pepys and Dyse: A Historical Note. Annals of Science 13, 137–147.

Fagnani, F., Zampieri, S., 2008. Randomized consensus algorithms over large scale networks. IEEE Journal on Selected Areas in Communications 26, 634–649.

Flajolet, P., Martin, G.N., 1985. Probabilistic counting algorithms for data base applications. Journal of Computer and System Sciences 31, 182–209.

Garin, F., Schenato, L., 2011. A survey on distributed estimation and control applications using linear consensus algorithms. Springer. volume 406. chapter 3. pp. 75–107.

Lehmer, D.H., 1955. The distribution of totatives. Canadian Journal of Mathematics 7, 347–357.

Massouliè, L., Merrer, E.L., Kermarrec, A.M., Ganesh, A., 2006. Peer counting and sampling in overlay networks: random walk methods, in: Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing, pp. 123–132.

Pepys, S., 1929. Private correspondence and miscellaneous papers of Samuel Pepys, 1679–1703: in the possession of J. Pepys Cockerell. Published for G. Bell and sons, ltd., edited by J. R. Tanner.

Rubin, E., Evans, F.B., 1961. On "Pepys, Newton, and Bernoulli probability" by Emil Schell. Reader observations on recent discussions, in the series "Questions and Answers". The American Statistician 15, 29–30.

Rubin, E., Schell, E.D., 1960. Questions and Answers on "Samuel Pepys, Isaac Newton, and Probability". The American Statistician 14, 27–30.

Sirken, M., Shimizu, I., 1999. Population based establishment sample surveys: The Horvitz-Thompson estimator. Survey Methodology 25, 187–191.

Stigler, S.M., 2006. Isaac Newton as a Probabilist. Statistical Science 21, 400–403.

Turnbull, H.W., 1961. The Correspondence of Isaac Newton, vol. 3, 1688-1694. Published for the Royal Society, edited by Cambridge University Press.

Xiao, L., Boyd, S., Kim, S.J., 2007. Distributed average consensus with least-mean-square deviation. Journal of Parallel and Distributed Computing 67, 33–46.