The Sum of Weighted Balls

Jop F. Sibeyn

RUU-CS-91-37
October 1991

Utrecht University

Department of Computer Science

Padualaan 14, P.O. Box 80.089,
3508 TB Utrecht, The Netherlands,
Tel. : ... + 31 - 30 - 531454

The Sum of Weighted Balls

Jop F. Sibeyn

Technical Report RUU-CS-91-37
October 1991

Department of Computer Science
Utrecht University
P.O.Box 80.089
3508 TB Utrecht
The Netherlands

ISSN: 0924–3275

The Sum of Weighted Balls

Jop F. Sibeyn *
Department of Computer Science, Utrecht University
P.O. Box 80.089, 3508 TB Utrecht, the Netherlands.
Email: jopsi@cs.ruu.nl

Abstract

In this paper we consider the problem of randomly distributing balls over n holes. The weight of the balls is given by a random variable. This problem models several problems from computer science, most importantly randomized load balancing. We analyse the distribution of the sums of the weights of the balls in the holes. We will carry out an analysis using the Chernoff bound and the Hoeffding inequality. Often $\mathcal{O}(n \cdot \log n)$ balls are enough to assure good balancing of the weights. If the expected weight of the balls is much smaller than their maximal weight, it is very useful to gather balls to super-balls. We also give an analysis based on the moment-generating function of the ball weights. We derive the moment-generating function of the probability-mass function of the sum of weights. From this moment-generating function it is easy to derive moments and variance.

Keywords

Probability theory, binomial distribution, Chernoff bound, Hoeffding inequality, moment-generating function, moments, variance, load balancing, random graphs.

1 Introduction

In this paper we analyse a problem from probability theory with several applications in computer science. The problem concerns the random distribution of r balls over n holes. The weights of the balls are given by some random variable X. The question is how large r must be in order to assure with high probability that all the sums of the weights of the balls lie sufficiently close to the expected value and thus close to each other. This minimal r depends on parameters of X, such as the size of its support, its expectation, and the allowed spread of the sums. If X is a constant random variable, all the balls will have the same weight and the only thing that matters is how many balls will end up in a hole. This problem is an instance of the coupon-collector-problem, for which it is easy to show that $n \cdot \log n$ balls are enough for bounding the sums of the weights of the balls within a small constant factor. Allowing the weighs of the balls to be randomly distributed makes the problem much more interesting from a mathematical point of view. However, the problem does not only deserve mathematical interest: In this form it models a problem for both random graphs and randomized load balancing. These problems are:

- What is the sum of the capacities of all edges leaving a vertex of a directed random graph?

*The work of the author was financially supported by the Foundation for Computer Science (SION) of the Netherlands Organization for Scientific Research (NWO). This research was partially supported be the ESPRIT II Basic Research Actions Project of the EC under contract No. 3075 (Project ALCOM). This paper was written during a stay of the author at the Hebrew University in Jeruzalem.

- What is the sum of the lengths of the tasks scheduled to a PU in a parallel network with random scheduling?

Of course, these questions cannot be answered. The question we will answer is how many balls (edges, tasks) we need to assure that the sums under consideration differ by at most a constant factor.

The number of balls we need to assure good balancing of the sums of the weights depends on the distribution of X. We will consider special cases (binomial, exponential and normal distribution); and the most general case, that we only know the support of the distribution function of X. The more we know of X, the sharper we can bound the required number of balls. For some applications, e.g., the second problem, it might be desirable to reduce the number of balls that are actually distributed. This happens when there is overhead involved in every distribution operation. This means that we would like to gather some balls to one super-ball. It turns out that especially the technique of gathering balls until the sum of their weights exceeds a threshold value is very good. Here we mentioned just two apparent applications but the problem of distributing balls with weight according to a random variable may appear whenever a randomized algorithm is given for a problem in which weights play a role.

In an earlier paper [7] we analysed a maxflow algorithm that combined some observations on the character of random graphs, with a load-balancing algorithm. We used the Chernoff bound and the Hoeffding inequality to give an estimate of the difference of the sum of all weights of the edges coming into a vertex and the sum of the weights of all edges leaving a vertex (the first problem). This gave pretty sharp estimates with which we could derive the desired results. These results were given in terms of: "With extremely high probability the calculation will not take more time than ...". This conforms with the usual way to express the results of ana;yses of average-case algorithms. Typically in this kind of analysis, there is made only a distinction between events that may occur and those that occur sufficiently rare to consider them as non-occurring. This is rather coarse an approach. More informative, and much tougher to obtain, would be results like: "The probability-mass function (**pmf**) of the calculation time t of the algorithm is given by $f(t)$". It seems unlikely that for any algorithm of interesting complexity this can be achieved. Nevertheless, it is at least informative, and may even turn out to give better estimates, to carry out an analysis along these lines so far as possible. Inspired by this we will try to give an alternative analysis of summing the weight of the balls that find their way to a hole. We cannot give an expression for the pmf of the sum but it is not too difficult to give its moment-generating function (**mgf**). In theory this gives an expression for the pmf of the sum but only in terms of an unsolvable integral. This is no serious limitation however: From the mgf of the sum we can derive expressions for the moments of the sum and in particular for the variance. This does not imply such strong bounds as given by the Chernoff bound or the Hoeffding inequality but it gives much better real understanding of what is happening.

The remainder of this paper is organized as follows: Section 2 is devoted to a presentation of known facts from probability theorem and arranging them in a suited form for application. We will give a little bit forgotten form of the Hoeffding inequality involving the mean of the distribution X of the ball weights. For distributions which are strongly asymmetric this gives much sharper estimates than the usual form of the Hoeffding inequality, involving only the support of X. In section 3 we will analyse the sums of weights of edges (balls) coming into a vertex (hole) using the Chernoff bound and the Hoeffding inequality. In particular we will address the question "how many balls there must be to assure that the sums of the ball weights differ by at most a given constant factor". The analysis is essentially the same as in [7] but is carried out in more detail. However, there is something new: The more involved version of the Hoeffding inequality can be applied here to show that the minimal degree a graph must have to assure that the total capacity of the in-edges differs at most a constant fraction from the capacity of the out-edges, is lower than previously estimated. Section 4 is devoted to an analysis of the spread of the sum of the ball weights as a function of the number of balls. In section 5 we will consider the use and problems of gathering balls to super-balls before distributing them. We distinguish two principal techniques: Gathering a fixed number of balls; gathering balls until the sum of their

weights exceeds a threshold. In section 6, we will carry out the alternative analysis of the problem involving pmf and mgf of the weights of the balls and their sum, resulting in the calculation of the variance of the sum.

2 Hoeffding inequality

In this section we introduce the Chernoff bound and the Hoeffding inequality. These are well-known bounds on the tail of the distribution of a sum of random variables. Both bounds can be found in many standard works on probability theory. Our basic formulation of the Chernoff bounds were taken from Hagerup and Rüb [3]. The Hoeffding inequality we took from Hofri [6] and will be derived in appendix B. Often we will have a pair of formulas of the form $P(A \geq f(1+\epsilon) \cdot B) \leq C$, $P(A \leq f(1-\epsilon) \cdot B) \leq C$, with A, B and C expressions and f a (simple) function. Such a pair of formulas will be presented concisely with the following notation: $P(A \succeq f(1+\epsilon) \cdot B) \leq C$.

A random variable B is a Bernoulli trial with parameter p if $P(B=0) = 1-p; P(B=1) = p$. We consider n independent Bernoulli trials B_i with parameter p_i and their sum $S_n = \sum_i B_i$. Let $p = \sum_i p_i/n$. The Chernoff bound gives

$$P(S_n \geq (1+\epsilon) \cdot n \cdot p) \leq e^{-\epsilon^2 \cdot n \cdot p/3},\ 0 \leq \epsilon \leq 1, \tag{1}$$

$$P(S_n \leq (1-\epsilon) \cdot n \cdot p) \leq e^{-\epsilon^2 \cdot n \cdot p/2},\ 0 \leq \epsilon \leq 1. \tag{2}$$

If the B_i are equally distributed with parameter p, then $S_n = S_{n,p}$, the binomial distribution with parameter n and p. (1) and (2) are mostly used in this form.

More generally, we want to consider the sum Z_n of n mutually independent random variables X_i with mean $\mu_i = E[X_i]$. Let $\mu = \sum_i \mu_i/n$. Sharp bounds on the tail probabilities are given by the Hoeffding inequalities. First we give some results for the case that $0 \leq X_i \leq 1$ for all i. The strongest form of the Hoeffding inequality is [6]

$$P(Z_n \geq n \cdot (\mu+t)) \leq \left[(\mu/(\mu+t))^{\mu+t} \cdot ((1-\mu)/(1-\mu-t))^{1-\mu-t}\right]^n \tag{3}$$

Using an estimate from [3], we find from (3)

$$\begin{aligned} P(Z_n \succeq (1+\epsilon) \cdot n \cdot \mu) &\leq \left[\frac{1}{(1+\epsilon)^{1+\epsilon}} \cdot \left(\frac{1-\mu}{1-(1+\epsilon)\cdot\mu}\right)^{1/\mu-(1+\epsilon)}\right]^{n\cdot\mu} \\ &\leq \left[\frac{e^\epsilon}{(1+\epsilon)^{1+\epsilon}}\right]^{n\cdot\mu} \leq 1.2 \cdot e^{-\epsilon^2 \cdot n \cdot \mu/2},\ 0 \leq \epsilon \leq 1. \end{aligned} \tag{4}$$

The generalizations of (4) to X_i with arbitrary support $[m_i, M_i]$ become particularly nice when the X_i are equally distributed. So, assume $m \leq X_i \leq M$, for some m and M. In this case

$$\begin{aligned} P(Z_n \succeq (1+\epsilon) \cdot n \cdot \mu) &= P(\sum_i (X_i - m)/(M-m) \succeq ((1+\epsilon)\cdot\mu - m) \cdot n/(M-m)) \\ &= P(\sum_i X_i' \succeq (1+\epsilon') \cdot n \cdot \mu'), \end{aligned}$$

where $X_i' = (X_i - m)/(M-m)$, $\epsilon' = \epsilon \cdot \mu/(\mu - m)$ and $\mu' = (\mu - m)/(M-m)$. Substituting in (4) gives

$$P(Z_n \succeq (1+\epsilon) \cdot n \cdot \mu) \leq 1.2 \cdot e^{-\epsilon^2 \cdot n \cdot \mu^2/(2\cdot(\mu-m)\cdot(M-m))}. \tag{5}$$

An important special case is $m = 0$. Then,

$$P(Z_n \succeq (1+\epsilon) \cdot \mu \cdot n) \leq 1.2 \cdot e^{-\epsilon^2 \cdot n \cdot \mu/(2\cdot M)}. \tag{6}$$

Noticing that $\mu^2/(\mu - m)$ has minimal value $4 \cdot m$ (for $\mu = 2 \cdot m$), we can still use (5) in case μ is an unknown value of the interval $[m, M]$. The least we can say then, is

$$P(Z_n \succeq (1+\epsilon) \cdot n \cdot \mu) \leq 1.2 \cdot e^{-2\cdot\epsilon^2 \cdot n \cdot m/(M-m)}. \tag{7}$$

We assumed that $0 \notin [m, M]$. Otherwise we cannot exclude $\mu = 0$ and there is no direct way to apply the Hoeffding inequality in this form to Z_n[1].

The four types of relations we introduced in this section will turn out to be just what we need in the derivations of the following sections.

3 Using Hoeffdings inequality

We consider the problem of distributing balls over holes. The weight of the balls is not fixed but is a random variable. We will analyse the sum of the weights of the balls in every hole. The problem has the following parameters:

- there are n holes;
- there are $r = \alpha \cdot n$ balls;
- balls have weight according to a random variable X, with $E[X] = \mu$.

Define an event a to occur with very high probability (**vhp**) if $P(\neg a) < n^{-\omega}$, for some constant $\omega > 0$. The goal of this section is to determine for every $0 < \epsilon \leq 0.5$ the smallest α that assures that the weights of the balls in the holes differ by at most a factor $(1+\epsilon)^2/(1-\epsilon)^2$ vhp.

We set $q = 1 - p$. Let k_i be the discrete random variable giving the number of balls that end up in in hole i. k_i is binomially distributed[2] with parameters $r, 1/n$. Applying (1) and (2) gives

$$\begin{aligned} P(k_i \geq (1+\epsilon) \cdot \alpha) &\leq e^{-\epsilon^2 \cdot \alpha/3}, \\ P(k_i \leq (1-\sqrt{2/3} \cdot \epsilon) \cdot \alpha) &\leq e^{-\epsilon^2 \cdot \alpha/3}. \end{aligned}$$

By $\log n$ we will mean $^2\log n$. Then, $e^{\log n} > n^{1.4}$. Thus,

Lemma 1 *If $\alpha \geq 3 \cdot \log n / \epsilon^2$, then*

$$\begin{aligned} P(k_i \geq (1+\epsilon) \cdot \alpha) &\leq n^{-1.4}, \\ P(k_i \leq (1-\sqrt{2/3} \cdot \epsilon) \cdot \alpha) &\leq n^{-1.4}. \end{aligned}$$

Let $Z_{n,i}$ be random variable giving the sum of the weights of the balls of hole i. Define

$$\alpha_{min} = \alpha_{min}(\epsilon) = 3 \cdot \max\{1, (\mu - m) \cdot (M - m)/\mu^2\} \cdot \log n / \epsilon^2. \tag{8}$$

There are two special cases:

$$\begin{aligned} \alpha_{min} &= 3 \cdot \max\{1, M/\mu\} \cdot \log n / \epsilon^2, \text{ if } m = 0, \\ \alpha_{min} &= 3 \cdot \max\{1, (M-m)/(4 \cdot m)\} \cdot \log n / \epsilon^2, \text{ if } \mu \text{ is unknown.} \end{aligned} \tag{9}$$

Putting an upper bound on $Z_{n,i}$ is rather easy: If $\alpha \geq \alpha_{min}$, then by lemma 1 and (5)

$$\begin{aligned} P(Z_{n,i} \geq (1+\epsilon)^2 \cdot \alpha \cdot \mu) &= P(\sum_{j=1}^{k_i} X_j \geq (1+\epsilon)^2 \cdot \alpha \cdot \mu) \\ &\leq n^{-1.4} + P(\sum_{j=1}^{(1+\epsilon)\cdot\alpha} X_j \geq (1+\epsilon) \cdot (1+\epsilon) \cdot \alpha \cdot \mu) \\ &\leq n^{-1.4} + 1.2 \cdot e^{-\epsilon^2 \cdot (1+\epsilon) \cdot \alpha \cdot \mu^2/(2 \cdot (\mu - m) \cdot (M-m))} \leq 2 \cdot n^{-1.4}. \end{aligned} \tag{10}$$

[1] Going back to (3) there is a form that is useful if one is interested in the range of values of Z_n and not the ratio between smallest and largest value Z_n can assume.

[2] The random vector $(k_1, \ldots, k_n)$ has as compound pmf the multinomial distribution function. For any i, the marginal pmf of k_i is then binomial [8, p 95].

In order to derive a lower bound on $Z_{n,i}$ with the same values of α, we must be more precise. It is here that we need $\epsilon \leq 0.5$:

$$\begin{aligned} P(Z_{n,i} \leq (1-\epsilon)^2 \cdot \alpha \cdot \mu) &= P(\sum_{j=1}^{k_i} X_j \leq (1-\epsilon)^2 \cdot \alpha \cdot \mu) \\ &\leq n^{-1.4} + P(\sum_{j=1}^{(1-\sqrt{2/3}\cdot\epsilon)\cdot\alpha} X_j \leq (1-1.15\cdot\epsilon)\cdot(1-\sqrt{2/3}\cdot\epsilon)\cdot\alpha\cdot\mu) \\ &\leq n^{-1.4} + 1.2\cdot e^{-(1.15\cdot\epsilon)^2\cdot(1-\sqrt{2/3}\cdot\epsilon)\cdot\alpha\cdot\mu^2/(2\cdot(\mu-m)\cdot(M-m))} \leq 2\cdot n^{-1.4}. \end{aligned}$$

Resuming,

Lemma 2 *If $\alpha \geq \alpha_{min}$, then*

$$P(Z_{n,i} \geq (1+\epsilon)^2 \cdot \alpha \cdot \mu) \leq 2 \cdot n^{-1.4}.$$

The main result of this section follows easily from lemma 2:

Theorem 1 *If $r = \alpha \cdot n$, with $\alpha \geq \alpha_{min}$, then*

$$(1-\epsilon)^2 \cdot \alpha \cdot \mu < Z_{n,i} < (1+\epsilon)^2 \cdot \alpha \cdot \mu, \text{ for all } i \text{ at the same time, vhp.}$$

Proof: We prove that the complementary case does not occur (vhp): $P(Z_{n,i} \leq (1-\epsilon)^2 \cdot \alpha \cdot \mu$ or $Z_{n,i} \geq (1+\epsilon)^2 \cdot \alpha \cdot \mu$ for some $i) \leq n \cdot P(Z_{n,i} \leq (1-\epsilon)^2 \cdot \alpha \cdot \mu) + n \cdot P(Z_{n,i} \geq (1+\epsilon)^2 \cdot \alpha \cdot \mu) \leq 4 \cdot n^{-0.4}$. □

The following corollary clearly expresses that the balancing, achieved by randomly distributing the balls, is fairly good:

Corollary 1 *If $\alpha \geq \alpha_{min}$, then $\max_i\{Z_{n,i}\}/\min_i\{Z_{n,i}\} \leq (1+\epsilon)^2/(1-\epsilon)^2$, vhp.*

For $\epsilon = 0.5$ the given factor is 9. For smaller ϵ we may approximate $(1+\epsilon)^2/(1-\epsilon)^2 = 1+4\cdot\epsilon$. In the case of load balancing (section 5) we may be more interested in the performance ratio of our algorithm. This means the value of $\max_i\{Z_{n,i}\}/Z_{n,\,opt}$, where $Z_{n,\,opt}$ is the maximum of the sums of the ball weights after optimally distributing them. We have

Corollary 2 *If $\alpha \geq \alpha_{min}$, then $\max_i\{Z_{n,i}\}/Z_{n,\,opt} \leq (1+\epsilon)^2$, vhp.*

Proof: Lemma 1 gives $Z_{n,\,opt} \geq \sum_j X_j/n \geq (1-\delta)\cdot\alpha\cdot\mu$, for all $\delta > 0$, vhp. On the other hand, being a little more careful in the derivation of (10) we find for a sufficiently small δ $P(Z_{n,i} \geq (1+\epsilon)^2/(1-\delta)\cdot\alpha\cdot\mu) \leq 2\cdot n^{-1.4}$, for all $\alpha \geq \alpha_{min}$. □

Some remarks: Implicitly we everywhere assumed that $\mu \neq 0$. One may think this is a restriction. However, if $\mu = 0$, we cannot expect to find a constant ratio between $\max_i\{Z_{n,i}\}$ and $\min_i\{Z_{n,i}\}$. In practical cases one often has $(M-m)/\mu = \mathcal{O}(1)$ in that case $\alpha_{min} = \mathcal{O}(\log n/\epsilon^2)$. Unbounded X are a problem for application of theorem 1. Often, however, it is easy to indicate m, M such that $m \leq \min_j\{X_j\} \leq \max_j\{X_j\} \leq M$ vhp. A good example of this is $X_j = \text{Exp}(\lambda)$ (appendix A). Although the X_j are unbounded, we can take $M = \log n/\lambda$. The corresponding value of $\alpha_{min} = 3\cdot\log^2 n/\epsilon^2$.

4 Spread of the sums

Until now we were satisfied if the sums of the ball weights differed by at most a constant fraction. Theorem 1 gives a lower bound on the number of balls that should be distributed to assure this. In this section we will

analyse how large the spread around the expected value is as a function of the number of balls. The results of this section have practical importance: Often, the number of tasks, balls, etc., will be given (or can be estimated) and we want to get an impression of the balancing.

Instead of starting with an ϵ and finding the minimal α that will assure that $(1-\epsilon)^2 \cdot \alpha \cdot \mu \leq Z_{n,i} \leq (1+\epsilon)^2 \cdot \alpha \cdot \mu$, for all i at the same time, we will give here the ϵ corresponding to a given α. From (8) we get a sharp bound: $\epsilon = [3 \cdot \max\{1, (\mu - m) \cdot (M - m)/\mu^2\} \cdot \log n / \alpha]^{1/2}$. For all ϵ, $1 - 2 \cdot \epsilon \leq (1-\epsilon)^2$ and for $\epsilon \leq 0.5$, $(1+\epsilon)^2 \leq 1 + 2.5 \cdot \epsilon$. Let $h(\alpha) = 2.5 \cdot \epsilon \cdot \alpha\mu$:

$$h(\alpha) = [19 \cdot \max\{1, (\mu - m) \cdot (M - m)/\mu^2\} \cdot \log n]^{1/2} \cdot \sqrt{\alpha} \cdot \mu. \tag{11}$$

Lemma 3 *If $\alpha \geq \alpha_{min}(0.5)$, then*

$$\alpha \cdot \mu - h(\alpha) \leq Z_{n,i} \leq \alpha \cdot \mu + h(\alpha), \text{ for all } i \text{ at the same time, vhp.}$$

Lemma 3 expresses clearly that the spread in the sum of the ball weights only grows with the square root of α, while the expectation of this sum grows with α itself.

It should be stated that the bounds we derived in this section and previously hold with very high probability. So, in exceptional cases, they may be violated. On the other hand, most sums will lie much closer to the expected value than suggested by these bounds. This normal behavior is expressed better by the standard deviation, σ. In section 6 we will see that σ is much smaller than the spread of lemma 3. That most sums lie within $\mathcal{O}(\sigma)$ from the expectation does not mean that there is a reasonable probability that all n sums will lie that close to the expected value. Just to assure that all sums lie between $\alpha \cdot \mu - h'$ and $\alpha \cdot \mu + h'$ with probability, say $1/2$, requires an h' of the same order as h. This is fundamental: The whole idea of using vhp bounding arises because often a final result is obtained as a conjunction of a polynomial number of cases. So, the individual cases must have had inverse polynomial probabilities. In such cases, it will not require much extra to bound the probability on the conjunction to inverse polynomial. In some way, the vhp is got for free.

5 Combining balls

In this section we will analyse the use of combining balls before distributing them. Previously [7] the sum of the weights of balls modeled the sum of the weights of the edges incident on a vertex in a random graph. The sum of the weights of balls can also be seen as a model for the total task length a scheduler schedules to every processing unit (**PU**) in a n PU parallel processing network. In the former case the randomness comes from assumption. In the latter case the randomness has to be brought in by using a random generator for allocating the tasks. In figure 1 we give a schematic representation of this random allocation of tasks to PUs. Producing many random numbers may be time consuming. Furthermore, scheduling a task to a PU will produce additional code and require the PU to interrupt its execution for accepting the task and storing it in a convenient way. All this brings about a lot of over-head for every task scheduled. It seems desirable to consider techniques to reduce this. Two natural strategies appear: The scheduler can

- gather k tasks before scheduling them as a block to a single PU;
- gather tasks together until the sum of their lengths exceeds a threshold $\tilde{m}$.

It will turn out that if the number of tasks to be scheduled is larger than the minimal number given by theorem 1, both ideas are very useful for reducing the relative amount of time PUs spend on overhead.

5.1 Gathering a fixed number of tasks

Let the lengths of the tasks be given by the random variable X, with $\mu = E[X]$. $Z_{n,i}$ is the sum of the lengths of the tasks that are allocated to PU i. The minimal number of tasks, that must be scheduled in order to assure

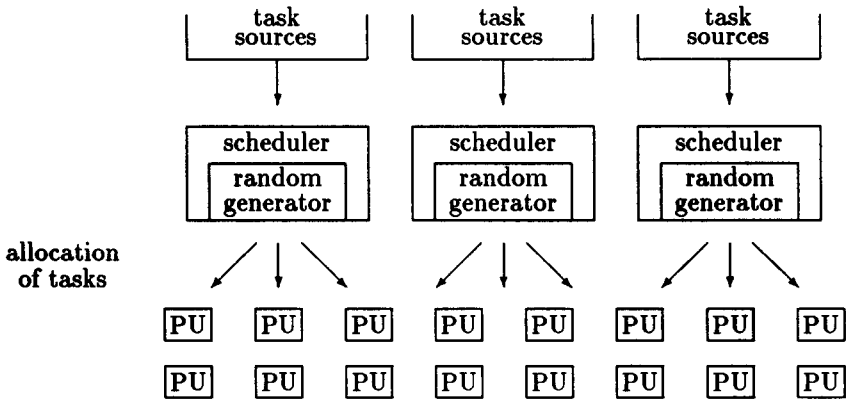

Figure 1: Random allocation of tasks to PUs.

$\max_i\{Z_{n,i}\}/\min_i\{Z_{n,i}\} \leq (1+\epsilon)^2/(1-\epsilon)^2$ is given by theorem 1. Suppose now, that the schedulers gather k tasks to one super-task. For given k, how many super-tasks must be scheduled to assure an $(1+\epsilon)^2/(1-\epsilon)^2$ ratio? What is a good choice of k?

Let $\tilde{X} = \sum_{j=1}^{k} X_j$ be the random variable giving the length of the super-tasks. We have $\tilde{\mu} = E[\tilde{X}] = k \cdot \mu$. Lemma 2 gives that for $k \geq \alpha_{min}(\tilde{\epsilon})$, $P(\tilde{X}_i \geq (1+\tilde{\epsilon})^2 \cdot \alpha \cdot \mu) \leq 2 \cdot n^{-1.4}$. This implies

Lemma 4 *If $k \geq \alpha_{min}(\tilde{\epsilon})$, then*

$$(1-\tilde{\epsilon})^2 \cdot k \cdot \mu \leq \tilde{X}_i \leq (1+\tilde{\epsilon})^2 \cdot k \cdot \mu, \text{ for all } i, \text{ vhp.}$$

Thus, for the super-tasks we have the following parameters:

$$\tilde{m} = (1-\tilde{\epsilon})^2 \cdot k \cdot m, \ \tilde{\mu} = k \cdot m, \ \tilde{M} = (1+\tilde{\epsilon})^2 \cdot M.$$

By theorem 1 this gives $\tilde{\alpha}_{min}(\epsilon) = 3 \cdot \max\{1, 4 \cdot \tilde{\epsilon}^2 \cdot (2-\tilde{\epsilon}^2)\} \cdot \log n / \epsilon^2$, for the minimal number of super-tasks divided by n that must be scheduled in order to get a ratio $(1+\epsilon)^2/(1-\epsilon)^2$ of the loads of the PUs. This number is minimal if $\tilde{\epsilon}^2 = 1 - \sqrt{3}/2$. We then have to schedule only $3 \cdot n \cdot \log n / \epsilon^2$ super-tasks. For the corresponding k we can take $k = \alpha_{min}(1-\sqrt{3}/2) < 23 \cdot (\mu - m) \cdot (M-m)/\mu^2 \cdot \log n$. We showed

Theorem 2 *If $r = \alpha \cdot n$, with $\alpha \geq 23 \cdot \log n \cdot \alpha_{min}$, then gathering $k = \alpha \cdot \epsilon^2/(3 \cdot \log n)$ tasks to a super-task before scheduling them, reduces the number of scheduling actions to $3 \cdot n \cdot \log n / \epsilon^2$ and gives*

$$(1-\epsilon)^2 \cdot \alpha \cdot \mu < Z_{n,i} < (1+\epsilon)^2 \cdot \alpha \cdot \mu, \text{ for all } i \text{ at the same time, vhp.}$$

Theorem 2 assumes that the total number of tasks to schedule r is known. It is more realistical to assume that it is known that r is large enough but not how large. In that case one could take $k = 23 \cdot (\mu - m) \cdot (M-m)/\mu^2 \cdot \log n$. This kills the bad effect of very unbalanced distributions X without requiring much more tasks to schedule.

5.2 Gathering until a threshold is exceeded

In section 5.1 we gathered a fixed number of tasks to a super-task. It seems better to gather tasks until the sum of their lengths exceeds a threshold value $\tilde{m}$. This approach can only be carried out if the scheduler can obtain in $\mathcal{O}(1)$ time a reliable bound on on the calculation time of every task. This seems to require that this information is determined by the task generators and attached to the tasks. A smaller disadvantage over the approach of packing a fixed number of tasks before scheduling them, is that the scheduler must sum integers

instead of just counting until k. As it turned out that the simpler approach of gathering a fixed number of tasks gives almost everything we could hope to find (theorem 2), it does not seem worth to spend time on the analysis of a second gathering technique. However, there is a small gap that could be bridged still: theorem 2 gives a lower bound on r that is a factor $23 \cdot \log n$ higher than the lower bound on r given by theorem 1. In this section we will show that this gap can be bridged except for a constant. Without loss of generality we will assume that $\mu = E[X] > 0$.

Tasks are taken together until the sum of their schedule lengths reaches or exceeds $\tilde{m}$. The random variable giving the length of the super tasks is denoted $\tilde{X}$. We suppose $\tilde{m} = \beta \cdot M$, for some $\beta > 0$. Obviously, $\tilde{m} = \beta \cdot M \leq \tilde{X} < (\beta + 1) \cdot M = \tilde{M}$. We will determine a suitable value for β. We do not know the value of $\tilde{\mu} = E[\tilde{X}]$ but, as explained at (7), we may assume[3] $\tilde{\mu} = 2 \cdot \tilde{m}$. Altogether,

$$\tilde{m} = \beta \cdot M,\ \tilde{\mu} = 2 \cdot \beta \cdot M,\ \tilde{M} = (\beta + 1) \cdot M.$$

$\tilde{\alpha}_{min}(\epsilon)$, the minimal number of super-tasks that must be scheduled in order to assure $max_i\{\tilde{X}_i\}/\min\{\tilde{X}_i\} \leq (1+\epsilon)^2/(1-\epsilon)^2$ vhp, is given by (9): $\tilde{\alpha}_{min} = 3 \cdot \max\{1, (\tilde{M}-\tilde{m})/4 \cdot \tilde{m}\} \cdot \log n\,/\epsilon^2 = 3 \cdot \max\{1, 1/(4 \cdot \beta)\} \cdot \log n\,/\epsilon^2$. Now, we get

Lemma 5 *For $\beta \geq 1/4$, $\tilde{\alpha}_{min} = 3 \cdot \log n\,/\epsilon^2$.*

We find it surprising that for such small values of β the super-tasks are already so well balanced. How many tasks there must be to assure that there are $\tilde{r} = \tilde{\alpha} \cdot n$ super-tasks? One expects this to be something like $\tilde{r} \cdot (M + \mu)/\mu$. It is difficult to give a precise expression. However, a good estimate is easy:

Lemma 6 $(1 - \delta) \cdot r \cdot \mu/\tilde{M} \leq \tilde{r} \leq (1 + \delta) \cdot r \cdot \mu/\tilde{m}$, *for any constant $\delta > 0$, vhp.*

Proof: No super-task has weight larger than $\tilde{M}$. Thus, using (5), $P(\tilde{r} < (1 - \delta) \cdot r \cdot \mu/\tilde{M}) \leq P(\sum_{i=1}^{r} X_i < (1 - \delta) \cdot r \cdot \mu) \leq 1.2 \cdot e^{-\delta^2 \cdot r \cdot \mu^2/(2 \cdot (\mu - m) \cdot (M - m))}$. $r > n$ is much larger than needed. Analogously: No super-task has weight smaller than $\tilde{m}$. Thus $P(\tilde{r} > (1 + \epsilon) \cdot r \cdot \mu/\tilde{m}) \leq P(\sum_{i=1}^{r} X_i > (1 + \epsilon) \cdot r \cdot \mu)$. For the rest we get the same derivation. □

Lemma 5 and lemma 6 are combined to prove

Theorem 3 *If $r = \alpha \cdot n$, with $\alpha \geq 6 \cdot \log n \cdot M/(\mu \cdot \epsilon^2)$, then gathering tasks to super-tasks until the sum of their weights exceeds $\beta \cdot M$, with $\beta = \alpha \cdot \mu \cdot \epsilon^2/(3 \cdot \log n \cdot M) - 1$, reduces the number of scheduling actions to $6.01 \cdot n \cdot \log n\,/\epsilon^2$, and gives*

$$(1 - \epsilon)^2 \cdot \alpha \cdot \mu < Z_{n,i} < (1 + \epsilon)^2 \cdot \alpha \cdot \mu,\ \textit{for all } i \textit{ at the same time, vhp.}$$

Proof: This choice of β gives $\tilde{M} = (\beta+1) \cdot M = \alpha \cdot \mu \cdot \epsilon^2/(3 \cdot \log n)$. Thus, by lemma 6, $\tilde{\alpha} \geq (1-\delta) \cdot \alpha \cdot \mu/\tilde{M} = (1 - \delta) \cdot 3 \cdot \log n\,/\epsilon^2$, for any $\delta > 0$. This is slightly smaller than $\tilde{\alpha}_{min}$, of lemma 5. Going back to lemma 1 we see that this is no problem. Lemma 6 also gives $\tilde{\alpha} \leq (1 + \delta) \cdot 6 \cdot \log n\,/\epsilon^2$. □

Theorem 3 gives us the optimal reduction of actual scheduling operations without a requirement for more tasks to schedule. This in contrast to the technique of gathering a fixed number of tasks to a super-task. Reduction of the number of tasks to distribute to $\mathcal{O}(n \cdot \log n)$ can only be realized in the ideal situation that the total number of tasks to schedule is known. If we do not know r in advance, then the best thing we can do is to take some $\beta = \Theta(1)$. E.g., simply $\beta = 1$. This eliminates all small tasks and reduces the number of times a scheduler is used by a factor M/μ.

[3] For $\beta \geq 1$, $2 \cdot \tilde{m} = 2 \cdot \beta \cdot M \geq (1 + \beta) \cdot M = \tilde{M}$. So, if $\beta \geq 1$, the worst possible value for $\mu \in [\tilde{m}, \tilde{M}]$ is $\tilde{M}$. It is clear that this is not the true value of $\tilde{\mu}$ (which will lie close to $\beta \cdot M + \mu$), but we can work with it.

6 Calculating the variance

In this section we analyse almost the same problem as in section 3. We consider the problem of shooting balls at a target. Not all balls hit the target and they do not all have a fixed weight. We will analyse the distribution of the sum of the weights of the balls that hit the target. The problem has the following parameters:

- r balls are shot at a target;
- hitting probability is p;
- balls have weight according to a random variable X.

In our earlier setting we had $p = 1/n$. In addition we assume that the distribution of X and its mgf $M_X(s)$ is known. Knowing more we can also derive more precise results on $Z_n = X_1 + X_2 + \cdots + X_N$, the random sum. E.g., we can calculate $\mathrm{Var}[Z_n]$. In notation and ideas we follow [8]. Let $p_N(k)$ be the probability that the discrete random variable N, giving the number of balls that hit the target, has value k. Generally one assumes that $p_N(k)$ is known. In our case, shooting r times at a target with hitting probability p, clearly N is binomial with parameter p. Analogously to [8, p 261] we get

$$M_{Z_n}(s) = \sum_{k=0}^{r} (M_X(s))^k \cdot p_N(k) = \sum_{k=0}^{r} (M_X(s))^k \cdot \left(\begin{array}{c} r \\ k \end{array} \right) \cdot p^k \cdot q^{r-k} = (M_X(s) \cdot p + q)^r \,. \tag{12}$$

This is very general. The only thing we have used is that the probability that the target is hit k times is given by the binomial distribution. Putting $t = i \cdot s$ (where $i = \sqrt{-1}$), we get the Fourier transform expression for Z_n:

$$N_{Z_n}(t) = M_{Z_n}(i \cdot s) = (N_X(t) \cdot p + q)^r \,.$$

In theory we could calculate from this the pmf of Z_n:

$$f_{Z_n}(x) = \frac{1}{2 \cdot \pi} \cdot \int_{-\infty}^{\infty} e^{-i \cdot x \cdot t} \cdot N_{Z_n}(t)\, \mathrm{d}t = \frac{1}{2 \cdot \pi} \cdot \int_{-\infty}^{\infty} e^{-i \cdot x \cdot t} \cdot (N_X(t) \cdot p + q)^r \, \mathrm{d}t.$$

However, this relation does not seem to be practical. (12) is more useful:

$$\begin{aligned} E[Z_n] &= \frac{\mathrm{d}}{\mathrm{d}s} M_{Z_n}(s)\,|_{s=0} \\ &= r \cdot (M_X(s) \cdot p + q)^{r-1} \cdot p \cdot \frac{\mathrm{d}}{\mathrm{d}s} M_X(s)\,|_{s=0}, \\ E[Z_n^2] &= \frac{\mathrm{d}^2}{\mathrm{d}s^2} M_{Z_n}(s)\,|_{s=0} \\ &= r \cdot (r-1) \cdot (M_X(s) \cdot p + q)^{r-2} \cdot p^2 \cdot \left(\frac{\mathrm{d}}{\mathrm{d}s} M_X(s) \right)^2 + r \cdot (M_X(s) \cdot p + q)^{r-1} \cdot p \cdot \frac{\mathrm{d}^2}{\mathrm{d}s^2} M_X(s)\,|_{s=0}. \end{aligned}$$

Using $M_X(0) = E[X^0] = 1$, $\frac{\mathrm{d}}{\mathrm{d}s} M_X(s)\,|_{s=0} = E[X]$ and $\frac{\mathrm{d}^2}{\mathrm{d}s^2} M_X(s)\,|_{s=0}$, these relations reduce to

$$\begin{aligned} E[Z_n] &= r \cdot p \cdot E[X], \\ E[Z_n^2] &= r \cdot (r-1) \cdot p^2 \cdot E^2[X] + r \cdot p \cdot E[X^2]. \end{aligned} \tag{13}$$

Now we get a nice relation for the variance of Z_n:

$$\mathrm{Var}[Z_n] = E[Z_n^2] - E^2[Z_n] = r \cdot p \cdot E[X^2] - r \cdot p^2 \cdot E[X^2] = r \cdot p \cdot (\mathrm{Var}[X] + q \cdot E^2[X]), \tag{14}$$

where $q = p - 1$. We find that $E[Z_n]$ and $\mathrm{Var}[Z_n]$ are given in terms of $E[X]$ and $\mathrm{Var}[X]$. So, the assumption at the beginning of this section that we know the distribution and mgf of X turns out to be superfluous. That both

$E[X]$ and $\mathrm{Var}[X]$ appear in the expression for $\mathrm{Var}[Z_n]$ is logical: There are two contributions to the uncertainty of the value of Z_n, the uncertainty in the number of contribuants and the uncertainty in the contribution of each of them. This fact is also reflected in our two step approach of the derivation of theorem 1. Both terms of (14) may dominate, depending on the value of q, $E[X]$, and $\mathrm{Var}[X]$. An important consequence of (14) is that the standard deviation of Z_n, $\sqrt{\mathrm{Var}[Z_n]}$, only grows with $\sqrt{r}$, whereas $E[Z_n]$ grows with r. Random variables with a well behaving distribution assume with high probability values within $\mathcal{O}(\sigma)$ from their expectation. Often Z_n will be well behaved. This means that often the spread of Z_n divided by the expectation of Z_n decreases with $\sqrt{r}$. This same fact appeared from lemma 3. Substituting $p = 1/n$ in (14), we can compare $h(\alpha)$ of (11) with $\sqrt{\mathrm{Var}[Z_n]} \simeq \sqrt{\alpha \cdot E[X^2]}$. Except for pathological X, $(\mu - m) \cdot (M - m)/\mu^2 \leq E[X^2]$. For such well behaving X we have $h(\alpha) = \Theta(\log^{1/2} n) \cdot \sqrt{\mathrm{Var}[Z_n]}$. This is a good result. In fact, the best possible: For large α, the $Z_{n,i}$ converge to normal distributions (appendix C). To bound a normally distributed random variable with mean μ and standard deviation σ between $\mu - h'$ and $\mu + h$ vhp, requires $h' = \Theta(\sqrt{\log n}) \cdot \sigma$ (corollary 4).

6.1 Examples

As examples we consider the exponential and the normal distribution. Basic properties of these distribution are derived in appendix A. Z_n is denoted $Z_{n,\lambda}$, $Z_{n,\mu,\sigma}$ respectively. Substituting in 13 and (14) gives

$$\begin{array}{rclcrcl} E[Z_{n,\lambda}] & = & r \cdot p/\lambda & & \mathrm{Var}[Z_{n,\lambda}] & = & r \cdot p \cdot (1+q)/\lambda^2, \\ E[Z_{n,\mu,\sigma}] & = & r \cdot p/\lambda & & \mathrm{Var}[Z_{n,\mu,\sigma}] & = & r \cdot p(\sigma^2 + q \cdot \mu^2). \end{array}$$

What do these relations imply in case $Z_{n,\lambda}$ and $Z_{n,\mu,\sigma}$ describe the sum of the lengths of the tasks scheduled to a PU? For $Z_{n,\lambda}$ it is rather easy: $\mathrm{Var}[Z_{n,\lambda}] \leq 2 \cdot r \cdot p/\lambda^2$. The variance is only a constant factor larger than when we would have known that exactly $r \cdot p$ balls would hit the target. This means that for the exponential distribution (and other distributions X with $E^2[X] = \mathcal{O}(\mathrm{Var}[X])$), randomly distributing the tasks does not perform less good than first collecting all tasks in one scheduler and then giving equal numbers of tasks to all PUs. For $Z_{n,\mu,\sigma}$, the situation is quite different: It seems that every relation between the two contribuants to $R_{\mu,\sigma}$ is possible, depending on μ and σ. This is not true: $N(\mu,\sigma)$ can only give a good description of the task lengths if $N(\mu,\sigma) > 0$ vhp. As, for $\mu \geq 0$, $P(N(\mu,\sigma) \leq 0) = \Theta(e^{-\mu^2/\sigma^2})$, this implies $\mu = \Omega(\sigma \cdot \sqrt{\log n})$. Together with $q = 1 - 1/n$, this gives $\mathrm{Var}[Z_{n,\mu,\sigma}] \simeq r \cdot p \cdot q \cdot \mu^2$: the variance is complete due to the uncertainty in the number of tasks a PU receives.

If we consider again the techniques of gathering tasks to super-tasks (cf. section 5), then it appears that gathering until a threshold focuses more strongly on getting super-tasks with small variance than the technique of gathering a fixed number of tasks. It is attractive to speculate that gathering until a threshold will in particular perform well when $\mathrm{Var}[Z_n]$ is dominated by $\mathrm{Var}[X]$, i.e. when $E^2[X] = o(\mathrm{Var}[X])$. On the other hand, the techniques may be just as good when $\mathrm{Var}[Z_n]$ is mainly a result of the uncertainty in the number of tasks a PU gets, i.e. when $\mathrm{Var}[X] = o(E^2[X])$.

Acknowledgement

This work was written under permanent stimulation of Yosi Ben Asscher, Aviad Cohen and Assaf Schuster. It was Yosi who came with the question leading to this paper. Another version, more geared towards the balancing problem and less to the mathematics, will appear as a joint work. In addition Assaf and Yosi contributed much to my feeling well in Jeruzalem. Frank den Hollander showed me the easy proof of (3), as given in appendix B.

7 Conclusion

In this paper we analysed the behavior of the sum of N identical and mutually independent random variables X_k, for the case that the number of summands N is a random variable. This analysis may be a first step in the analysis of load-balancing algorithms.

References

[1] Chow, Y.S., H. Teicher, *Probability theory, independency, interchangeability, martingales, second edition*, Springer Verlag, New York inc., 1987.

[2] Feller, W., *An introduction to probability theory and its applications*, Vol. 1, Wiley, New York, 1950.

[3] Hagerup, T., C. Rüb, A guided tour of Chernoff Bounds, *Techn. Rep. A 88/03*, Dep. of Comp. Sc., Universität des Saarlandes, Saarbrücken, Germany, 1988. Also, *Inf. Proc. Lett.*, 33 (1990), 305-308.

[4] Hoeffding, W., On the distribution of the number of successes in independent trials, *Ann. of Math. Stat.*, 27 (1956).

[5] -, Probability inequalities for sums of bounded random variables, *Journ. American Statistical Assoc.*, 58 (1963), 13-50.

[6] Hofri, M., *Probabilistic analysis of algorithms on computing methodologies for computer algorithms performance evaluation*, Springer Verlag, New York inc, 1987.

[7] Sibeyn, J.F., A pseudo-polylog average time parallel maxflow algorithm, *Techn. Rep. RUU-CS-90-19*, Dep. of Comp. Sc., Utrecht University, Utrecht, the Netherlands, 1990.

[8] Trivedi, K.S., *Probability and statistics with reliability, queuing, and computer science applications*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1982.

A Exponential and normal distribution

In this section we derive some of the relevant properties of the exponential and normal distribution. More details can be found in any introductory text on probability theory [2, 8]. Both distributions are of utmost importance. The normal distribution because of the central limit theorem; the exponential distribution because it is a continuous version of the Poisson distribution and because it is an easy but non-trivial positive distribution. As such it might model the length of tasks that are scheduled to PUs. This is even more true for the hypo-exponential distribution, which is a generalization of the exponential distribution with more parameters. All results for the exponential distribution can be generalized to results for the hypo-exponential distribution.

The exponential distribution with parameter λ, Exp(λ), has the following pdf f:

$$f(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ \lambda \cdot e^{-\lambda \cdot x} & \text{for } x > 0. \end{cases}$$

Elementary integration gives all the results we need:

$$\begin{aligned}
P(\mathrm{Exp}(\lambda) \geq t) &= \int_t^\infty \lambda \cdot e^{-\lambda \cdot x}\, \mathrm{d}x &&= -e^{-\lambda \cdot x}\big]_t^\infty &&= e^{-\lambda \cdot t},\\
E[\mathrm{Exp}(\lambda)] &= \int_0^\infty x \cdot e^{-\lambda \cdot x}\, \mathrm{d}x &&= -x \cdot e^{-\lambda \cdot x} - e^{-\lambda \cdot x}/\lambda\big]_0^\infty &&= 1/\lambda,\\
E[\mathrm{Exp}^2(\lambda)] &= \int_0^\infty x^2 \cdot e^{-\lambda \cdot x}\, \mathrm{d}x &&= -x^2 \cdot e^{-\lambda \cdot x} - 2 \cdot x \cdot e^{-\lambda \cdot x}/\lambda - 2 \cdot e^{-\lambda \cdot x}/\lambda^2\big]_0^\infty &&= 2/\lambda^2,\\
\mathrm{Var}[\mathrm{Exp}(\lambda)] &= E[\mathrm{Exp}^2(\lambda)] - E^2[\mathrm{Exp}(\lambda)] &&= 2/\lambda^2 - 1/\lambda^2 &&= 1/\lambda^2.
\end{aligned}$$

Corollary 3 *If the X_j are exponentially distributed, then $X_j \leq 2 \cdot \log n/\lambda$, for n^2 trials at the same time, vhp.*

Proof: The first inequality gives $P(X_j > 2 \cdot \log n/\lambda) < e^{-2 \cdot \log n} < 1/n^{2.8}$. Thus, $P(\bigvee_{j=1}^{n^2} X_j > 2 \cdot \log n/\lambda) \leq n^{-0.8}$. □

The other distribution function we consider in this section is the normal distribution with parameters μ and σ. The pdf f of $N(\mu, \sigma)$ is given by

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-(x-\mu)^2/(2 \cdot \sigma^2)}.$$

Basic results are

$$\begin{aligned}
\int_{-\infty}^{\infty} \tfrac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-(x-\mu)^2/(2 \cdot \sigma^2)}\, \mathrm{d}x &= \int_{-\infty}^{\infty} \tfrac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-x^2/2}\, \mathrm{d}x &&= 1,\\
E[N(\mu, \sigma)] = \int_{-\infty}^{\infty} \tfrac{x}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-(x-\mu)^2/(2 \cdot \sigma^2)}\, \mathrm{d}x &= \int_{-\infty}^{\infty} \tfrac{x+\mu}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-x^2/(2 \cdot \sigma^2)}\, \mathrm{d}x &&= \mu.
\end{aligned}$$

The calculation that $E[N^2(\mu, \sigma)] = \sigma^2 + \mu^2$, and thus that $\mathrm{Var}[N(\mu, \sigma)] = \sigma^2$, is more involved ([8, 203]).

Lemma 7 *For all $h \geq 1$,*

$$P(N(\mu, \sigma) \geq \mu + \sigma \cdot h) = P(N(\mu, \sigma) \leq \mu - \sigma \cdot h) < e^{-h^2/2}/2.$$

Proof: By symmetry of the normal distribution only one inequality has to be checked: $P(N(\mu, \sigma) \geq \mu + \sigma \cdot h) = P(N(0, 1) \geq h) = \int_h^\infty e^{-x^2/2}\, \mathrm{d}x/\sqrt{2 \cdot \pi} < \sum_{i=0}^\infty e^{-(h+i/k)^2/2}/(k \cdot \sqrt{2 \cdot \pi}) < \frac{e^{-h^2/2}}{k \cdot \sqrt{2 \cdot \pi}} \cdot \sum_{i=0}^\infty e^{-h \cdot i/k} \leq \frac{e^{-h^2/2}}{k \cdot \sqrt{2 \cdot \pi}} \cdot \frac{1}{1 - e^{-h/k}} \leq \frac{e^{-h^2/2}}{\sqrt{k \cdot 2 \cdot \pi}} \cdot \frac{1}{1 - e^{-1/k}} < e^{-h^2/2}/2$. The first "<" holds for all $k > 0$, the last "<" for $k \geq 3$. □

This result can be improved but for most purposes it is sharp enough, easy to derive and easy to remember. Using lemma 7, we find an analogue of corollary 3:

Corollary 4 *Let $m = \mu - \sigma \cdot 2 \cdot \sqrt{\log n}$ and $M = \mu + \sigma \cdot 2 \cdot \sqrt{\log n}$. If the X_j are normally distributed, then $m \leq X_j \leq M$, for n^2 trials at the same time, vhp.*

B Proof of the Hoeffding inequality

In this section we will show how the general form of the Hoefding inequality (3) can be derived using only the Markov inequality abd elementary manipulations. This is illustrative in its own right, but also it makes clear how one could incorporate the variance of X in relations expressing the minimal r. In the derivation we will assume that the X_i are identically distributed with $E[X] = \mu$. The case that the average of the $E[X_i]$ is μ gives the same relation but is more difficult to derive and does not show interesting new techniques.

Let $Z_n = \sum_{i=1}^{n} X_i$. Then,

$$\begin{aligned} P(Z_n \geq (\mu+t)\cdot n) &= \text{(for all } \lambda > 0) \\ P(e^{\lambda \cdot Z_n} \geq e^{\lambda\cdot(\mu+t)\cdot n}) &\leq \text{(Markov inequality [2])} \\ E[e^{\lambda\cdot Z_n}]/e^{\lambda\cdot(\mu+t)\cdot n} &= \text{(independency of the } X_i\text{)} \\ \left[E[e^{\lambda\cdot X}]/e^{\lambda\cdot(\mu+t)}\right]^n &= \text{(independency of the } X_i\text{)} . \end{aligned} \tag{15}$$

For $X \leq 1$ we can develop the numerator as follows:

$$E[e^{\lambda\cdot X}] = \sum_{i\geq 0} \frac{\lambda^i}{i!}\cdot E[X^i] \leq 1 + \sum_{i\geq 1} \frac{\lambda^i}{i!}\cdot E[X] = 1 + (e^\lambda - 1)\cdot\mu. \tag{16}$$

Substituting this in (15) gives

$$P(Z_n \geq (\mu+t)\cdot n) \leq \inf_{\lambda\geq 0}\{\left[1 + (e^\lambda - 1)\cdot\mu/e^{\lambda\cdot(\mu+t)}\right]^n\}.$$

For which λ the infimum is assumed? After taking the derivative we have to solve $0 = \mu\cdot e^\lambda \cdot e^{\lambda\cdot(\mu+t)} - (1 + (e^\lambda - 1)\cdot\mu)\cdot(\mu+t)\cdot e^{\lambda\cdot(\mu+t)}$. This gives $e^\lambda = (1-\mu)\cdot(\mu+t)/(\mu\cdot(1-\mu-t))$. Thus, by rearranging factors, we find

$$P(Z_n \geq (\mu+t)\cdot n) \leq \left[\frac{1 + (1-\mu)\cdot(\mu+t)/(1-\mu-t) - \mu}{((1-\mu)\cdot(\mu+t)/(\mu\cdot(1-\mu-t)))^{\mu+t}}\right]^n = \left[\frac{\mu}{\mu+t}\right]^{\mu+t}\cdot\left[\frac{1-\mu-t}{1-\mu}\right]^{\mu+t-1}.$$

This is (3). When we would like to introduce the second moment of X as well, we could have used instead of (16) a more refined relation:

$$E[e^{\lambda\cdot X}] \leq 1 + \lambda\cdot E[X] + \sum_{i\geq 2}\frac{\lambda^i}{i!}\cdot E[X] = 1 + \lambda\cdot\mu + (e^\lambda - 1 - \lambda)\cdot E[X^2].$$

Going on in the same way, we will not find a nice expression for e^λ that can be substituted. By numerical techniques a value of λ can be calculated by equating the derivative to 0.

C Convergency to the normal distribution

Satisfaction of the Lindeberg condition [1, ch 9.1] is necessary and sufficient for a sum of random variables to converge to normal. For independent identically distributed random variables, the case we consider, the only requirement to the Lindeberg condition is that the variance of the random variables is bounded. This seems very promising. However, in addition to convergency to normal, we also need sufficiently fast convergency. The Berry-Esseen theorem [1, ch 9.1] quantifies this convergency speed. For values around the mean it gives sharp results, but not for the values far away from the mean in which we are interested. We could not find a theorem that quantifies nicely the convergency of the tail of the sum of random variables to normal. Nevertheless, for special distributions, it may be reasonable to assume that even the tails converge sufficiently fast. Under this assumption, which will be made throughout this section, most results of this paper can be improved.

Let $\mu = E[X]$, $\sigma = \sqrt{\mathrm{Var}[X]}$. If $Z_n = \sum_{i=1}^{n} X_i$ converges to a normal distribution, then this must be to the normal distribution with parameters $n\cdot\mu$ and[4] $\sqrt{n}\cdot\sigma$. This is equivalent to saying that $Z'_n = (Z_n - n\cdot\mu)/(\sqrt{n}\cdot\sigma)$ converges to the standard normal distribution $N(0,1)$.

Lemma 8 *For sufficiently large n,*

$$P(Z_n \geq (1+\epsilon)\cdot n\cdot\mu) \leq e^{-n\cdot(\epsilon\cdot\mu/\sigma)^2/2}.$$

[4] It is a well-known fact that the variance of a sum of independent random variables is the sum of their variances. This is also a special case ($p = 1$) of (14).

Proof: For sufficiently large n, $P(Z_n \succeq (1+\epsilon)\cdot n\cdot\mu) \leq 2\cdot P(N(n\cdot\mu, \sqrt{n}\cdot\sigma) \succeq (1+\epsilon)\cdot n\cdot\mu)$. Lemma 7 with $h = \epsilon\cdot\sqrt{n}\cdot\mu/\sigma$ gives the result. □

Now, we consider again the problem of section 3. $Z_{n,i}$ is the random variable giving the sum of the weights of the balls that fall in hole i. Define

$$\alpha'_{min} = 3\cdot\max\{1, \sigma^2/\mu^2\}\cdot\log n/\epsilon^2. \tag{17}$$

Using lemma 1 and lemma 8, we can derive analogously to lemma 2:

Lemma 9 *If $\alpha \geq \alpha'_{min}$, then*

$$P(Z_{n,i} \succeq (1+\epsilon)^2\cdot\alpha\cdot\mu) \leq 2\cdot n^{-1.4}.$$

This gives

Theorem 4 *If $r = \alpha\cdot n$, with $\alpha \geq \alpha'_{min}$, then*

$$(1-\epsilon)^2\cdot\alpha\cdot\mu < Z_{n,i} < (1+\epsilon)^2\cdot\alpha\cdot\mu, \text{ for all } i \text{ at the same time, vhp.}$$

Going on in this way it turns out that we can replace all occurrencies of $(\mu - m)\cdot(M-m)/\mu^2$ in section 4 and section 5.1 by σ^2/μ^2. The results of section 5.2 are not improved if we may assume that the $Z_{n,i}$ are (almost) normally distributed: Under the technique of gathering until a threshold is exceeded, there is almost no relation between the distribution function of the super-tasks and the distribution function of the X_i.

It may happen in a practical situation that we have only limited knowledge of the X_i. E.g., we only know their support. In that case we will use the α_{min} of (9). The lack of knowledge prevents us from using (8) or even (17). It is interesting to investigate whether this matters a lot or not. The following examples show that for some X_i (8), (9) and (17), give large differences and that for some X_i they give the same value:

	$B_{m,2\cdot m}(1/2)$	$N(\mu,\sigma)$	$\mathrm{Exp}(\lambda)$
m	m	$\mu - 2\cdot\sqrt{\log n}$	0
M	$2\cdot m$	$\mu + 2\cdot\sqrt{\log n}$	$2\cdot\log n/\lambda$
μ	$3\cdot m/2$	μ	$1/\lambda$
σ	$m/2$	σ	$1/\lambda$
$(M-m)/(4\cdot m)$	1/4	$\sigma\cdot\sqrt{\log n}/(\mu - 2\cdot\sigma\cdot\sqrt{\log n})$	∞
$(\mu-m)\cdot(M-m)/\mu^2$	2/9	$8\cdot\log n\cdot\sigma^2/\mu^2$	$2\cdot\log n$
σ^2/μ^2	1/9	σ^2/μ^2	1

$B_{m,M}(p)$ is a generalized Bernoulli trial with parameter P: $P(B_{m,M}(p) = m) = 1-p$, $P(B_{m,M}(p) = M) = p$. $\mathrm{Exp}(\lambda)$ is the exponential distribution with parameter λ. The basic properties of $\mathrm{Exp}(\lambda)$ and $N(\mu,\sigma)$ are given in appendix A. If the X_i are normally distributed, then the $Z_{n,i}$ are normal as well [8, th 3.6.] and we are sure that the relations based on (17) can be used. We see that this gives a gain of a factor $8\cdot\log n$ over using (8).