On the sum of nonnegative independent random variables with unbounded variance

Uriel Feige
Department of Computer Science and Applied Mathematics
The Weizmann Institute
Rehovot 76100, Israel
feige@wisdom.weizmann.ac.il

September 5, 2003

Abstract

We prove the following inequality: for every positive integer n and every collection $X_1, \ldots, X_n$ of nonnegative independent random variables that each has expectation 1, the probability that their sum exceeds $n+1$ is at most $\alpha < 1$. Our proof produces a value of $\alpha = 12/13 \simeq 0.923$, but we conjecture that the inequality also holds with $\alpha = 1 - 1/e \simeq 0.632$.

1 A new inequality

For a random variable X, its typical value may be very different from its mean. In particular, the probability that X exceeds its mean may be arbitrarily close to 1. In some special case (e.g., when X is symmetric around its mean), the probability that X exceeds its mean is at most 1/2. The purpose of this manuscript is to investigate the probability that X exceeds its mean when X is the sum of n independent random variables. We show that under very general conditions, this probability is bounded away from 1, provided that we give ourselves a little slackness in exceeding the mean. Specifically, we prove the following inequality concerning the sum of independent nonnegative random variables.

Theorem 1 *For an arbitrary integer $n \geq 1$, let $X_1, \ldots, X_n$ be arbitrary nonnegative independent random variables, with expectations $\mu_1, \ldots, \mu_n$ respectively, where $\mu_i \leq 1$ for every i. Let $X = \sum_{i=1}^{n} X_i$, and let μ denote the expectation of X (hence, $\mu = \sum_{i=1}^{n} \mu_i$). Then for every $\delta > 0$,*

$$Pr[X < \mu + \delta] \geq \min[\delta/(1+\delta), 1/13] \tag{1}$$

The term $\delta/(1+\delta)$ in Theorem 1 is best possible, as one can take $X_1 = 1+\delta$ with probability $1/(1+\delta)$ and 0 otherwise, and all of the other X_i as the constant 1. This gives $\mu_i = 1$ for every i. For this case $Pr[X < \mu+\delta] = Pr[X_1 = 0] = \delta/(1+\delta)$. When δ is above some small constant (1/12 in the current statement of the theorem) the bounds in Theorem 1 are no longer tight. Moreover, for even larger δ (e.g., $\delta = 1$), it is no longer true that $Pr[X \leq \mu+\delta] \geq \delta/(1+\delta)$. One can take for every i, $X_i = n+\delta$ with probability $1/(n+\delta)$ and 0 otherwise. This gives $\mu_i = 1$ for every i, implying $\mu = n$. For this case $Pr[X < n+\delta] = (1-1/(n+\delta))^n$, which is roughly $1/e$ for large n.

It is our conjecture that for every value of δ and n, one of the two examples above is the worst case for $Pr[X < \mu+\delta]$. The conjecture, if true, would allow us to replace the constant 1/13 by $1/e$ in Theorem 1.

In may be instructive to consider how some standard probabilistic tools relate to Theorem 1. Consider first the case that the X_i are identically distributed. Then the central limit theorem implies that when n is large enough, X approaches the normal distribution and hence $Pr[X < \mu]$ approaches 1/2. However, in our Theorem 1 the variables X_i may depend on n, and hence n cannot be thought of as being "large enough" with respect to the X_i (even if they are i.i.d.). This relates to the fact that we place no bounds on the variance of the X_i, and hence standard bounds on deviations of random variables from their expectation (such as Chebyschev's bound, or Chernoff's bound) are not applicable.

The only restriction on the random variables (other than being independent) is their nonnegativity. In particular, this means that X is nonnegative, and that Markov's inequality can be used to show that $Pr[X \leq \mu+\delta] \geq \delta/(\mu+\delta)$. For the sum of independent identically distributed random variables, this bound tends to 0 as n grows (unlike the bound in Theorem 1). We note that when $\delta > \mu/12$, the bound in Theorem 1 does not improve over the bound obtainable from Markov's inequality. In fact, Markov's inequality is nearly best possible for lower bounding $Pr[X \leq \rho\mu]$ when ρ is large (say, $\rho > 10$). The new feature in Theorem 1 is in establishing that around the point $\rho = 1$, a slight increase of ρ (by $1/12n$, which is small when n is large) causes $Pr[X \leq \rho\mu]$ to jump from 0 to 1/13.

The author is aware of some work of nature similar to Theorem 1, though the proof techniques in these related works did not prove helpful for proving Theorem 1 (so far). There are certain results surveyed and developed by Siegel [3] that show that under certain conditions the median of the sum of random variables does not exceed the mean. This holds for example for the sum of Bernoulli random variables (if the mean is an integer). The book "How to gamble if you must" by Dubins and Savage [1] analyses strategies for gambling when the goal is to maximize the probability of ending up with a profit of δ. In typical scenarios, the strategies are allowed to be adaptive (next gamble may depend of outcomes of previous gambles) and the gambler is allowed to quit once a net profit of δ is achieved. One of the main findings of [1] is in establishing sufficient conditions under which the strategy of "playing boldly" is optimal. Informally, this strategy tries to reach a net profit of δ (taking into acount also previous losses) in one gamble. A typical example is the repeated doubling approach to gain one dollar when there are 50/50 odds, in which the gambler first gambles one dollar, and then doubles the gamble until the first win (or until he/she runs out of money). The scenario in Theorem 1 can be viewed as a version of "how to gamble in parallel". There are n gamblers who coordinate their gambling strategies in advance. Then, in parallel, they all enter different casino houses and cannot communicate with each other until they all leave the casino houses. The goal of the n gamblers is to reach a net profit of δ (in some standard units, such as millions of dollars). This needs to be done under the following additional restrictions: the amount of funds a gambler has on entering a casino house is at most 1, gamblers cannot borrow money once they are inside a casino house, and no casino house offers gambles on which the casino house looses in expectation. The correspondence with Theorem 1 is as follows. A variable X_i corresponds to the distribution of outcomes of the strategy of gambler i. The condition $\mu_i \leq 1$ follows from the limit on the amount of funds with which a gambler enters a casino house, and the fact that the casino house does not loose in expectation. The independence of the X_i corresponds to the gamblers not being able to communicate with each other once inside the casino houses. The outcome $X \geq \mu + \delta$ is the goal of the gamblers. Our results show that if $\delta \leq 1/12$ and the casino houses are unbiased (in expectation, the casino house also does not gain), the best policy of the gamblers is for all gamblers but one not to gamble at all, and have just one gambler try to reach a net profit of δ. This particular gambler may "play boldly". Perhaps more importantly, Theorem 1 implies that for every fixed $\delta > 0$, the probability of achieving a net profit of δ is bounded away from 1, regardless of the number of gamblers involved (as long as the restrictions listed above hold).

Theorem 1 can in principle be used whenever one is interested in bounding the probability that the sum of independent random variables significantly exceeds its expectation. However, in many of these cases the random variables are known to have some additional properties (e.g., bounded variance), and more useful results can be derived by other means. The application that motivated the development of the inequality (1) is described in a companion paper [2]. The purpose of the current paper is to bring this inequality to the attention of other researchers, who will hopefully find it both instructive and useful.

2 Proof of the main theorem

We our now ready to prove Theorem 1.

Proof: Fix n, δ, and arbitrary nonnegative random variables $X_1, \ldots, X_n$ with means at most 1. We prove that Equation (1) holds. We may assume that the support of every random variable is composed of a finite set of values. (This is a standard argument, but we sketch it for completeness. Any value larger than $\mu + \delta$

in the support of a random variable can be lowered to $\mu + \delta$, without increasing the probability that $X < \mu + \delta$. Thereafter, any continuous random variable can be approximated by a discrete random variable with the same mean and whose support includes only multiples of ϵ, where ϵ is chosen to be much smaller than δ/n. For these new random variables, $X'_1, \ldots, X'_n$, the event $X' < \mu + \delta'$ where $\delta' = \delta - \epsilon n$ implies that for the original variables, $X < \mu + \delta$. By making ϵ arbitrarily small, we can make δ' arbitrarily close to δ.)

Our proof of Equation (1) consists of a sequence of transformations on the variables X_i. For simplicity of notation, we keep calling the random variables by X_i, their sum by X and the expectation of X by μ, even though the random variables themselves and μ do change by the transformations. The invariant kept by the transformations is that $Pr[X < \mu + \delta]$ does not increase (though the interpretation of X and μ, but not δ, does change).

Our first transformation, called *remove constant*, is applied whenever there is a random variable that is constant. That is, we have a random variable X_i such the $Pr[X_i = \mu_i] = 1$. Such a random variable is removed, and μ is updated to $\mu - \mu_i$. Clearly, $Pr[X < \mu + \delta]$ does not change by *remove contant*.

Our next tranformation, called *reduce support*, is applied to every random variable whose support has at least three values, and replaces it with a new random variable with the same mean, and whose support includes at most two values from the original support.

Lemma 2 *Let X_i be a random variable whose support includes at least three values. Then X_i can be replaced by a new variable (which we shall also call X_i) without changing μ_i, and whose support includes only two values from the original support of X_i. This can be done without increasing $Pr[X < \mu + \delta]$.*

Proof: Let $\{v_1, \ldots, v_k\}$ be the support of the original X_i, and for $1 \leq j \leq k$, let q_j denote the conditional probability of the event $[X < \mu + \delta]$, conditioned on the event $[X_i = v_j]$. For the new X_i and for $1 \leq j \leq k$, we wish to select $p_j = Pr[X_i = v_j]$, under the restrictions that the mean of X_i is preserved, and that $Pr[X < \mu + \delta]$ does not increase. This can be expressed by the following linear program over the variables p_j:
Minimize $\sum_{j=1}^{k} q_j p_j$
subject to:

- $\sum_{j=1}^{k} p_j = 1$
- $\mu_i = \sum_{j=1}^{k} p_j v_j$
- $p_j \geq 0$, for every j.

The above linear program is feasible (as the probabilities associated with the original X_i satisfy the constraints). By the theory of linear programming, there is a *basic* optimal solution in which at most two p_j are nonzero. □

Our next transformation, called *align with 0*, is applied to every random variable whose support has two values and these values are greater than 0 (say X_i has value v_1 with probability p and v_2 with probability $(1-p)$, with $0 < v_1 < v_2$), and replaces it by a random variable that has value $v_1 - v_1 = 0$ with probability p, and has value $v_2 - v_1$ with probability $(1-p)$. This decreases μ_i by v_1, and to compensate for it we update μ to $\mu - v_1$. Clearly, $Pr[X < \mu + \delta]$ does not change by *align with 0*.

The next transformation that we describe, called *merge*, takes the two random variables with smallest mean, and merges them into one variable as follows. Assume without loss of generality that X_i and X_j have smallest mean, and that $X_i = v_1$ with probability p and 0 with probability $(1-p)$, and that $X_j = v_2$ with probability q and 0 with probability $(1-q)$. Then the merge operation replaces X_i and X_j by a new variable in three steps:

1. Replace X_i and X_j by a new random variable that is distributed like $X_i + X_j$. Namely it has value v_1 with probability $p(1-q)$, value v_2 with probability $q(1-p)$, value $v_1 + v_2$ with probability pq, and value 0 with probability $(1-p)(1-q)$.

2. Apply *reduce support* to this new random variable.

3. Apply *align with 0* or *remove constant* to the new random variable (if applicable).

It is easy to see that the transformation *merge* does not increase $Pr[X < \mu - \delta]$.

The sequence of transformations that we perform can be partitioned into two stages. By now we are ready to describe the first stage.

Stage 1:

1. Whenever possible, apply *remove constant*.

2. Apply *reduce support* until all random variables have support of size at most two. (Different variables may have different support.) This takes at most n steps.

3. Apply *align with 0* to all variables. This takes at most n steps.

4. Apply *merge* until either it cannot be applied anymore (because the number of random variables was reduced to one), or all random variables have mean at least 1/2 (whichever happens first). This takes at most $n-1$ steps, because with each application of *merge*, the number of random variables decreases.

Let $X_1, X_2, \ldots, X_{n'}$ be the random variables that we remain with when stage 1 ends. We assume that they are ordered in order of decreasing μ_i. Their number n' may be smaller than n, because some of the transformations remove random variables. These are not arbitrary random variables, as each of them has a support of two values, one of which is 0, and the stopping condition for the merge transformations has been reached. For random variables as above, we shall in fact prove that one of the following two alternatives must hold.

The two alternatives. At the end of stage 1, one of the following two alternatives must hold:

- $Pr[X \geq \mu + \delta] = 0$,
- $Pr[X < \mu] \geq \min[\delta/(3/2 + \delta), 1/13]$.

Note that in the second alternative considers the event $[X < \mu]$, which is even stronger than $[X < \mu + \delta]$. The first bound in the second alternative does not quite imply Equation (1), because we might have $\delta/(3/2 + \delta) \leq Pr[X < \mu] < \delta/(1 + \delta)$. Cases like this will be handled by showing that one merge operation before the end of stage 1 we had $Pr[X < \mu+\delta] \geq \delta/(1+\delta)$. (Here we will need to consider the original event $[X < \mu + \delta]$ rather than the stronger event $[X < \mu]$. For example, consider the case when $n = 2$, X_1 has support $(0, 1 + \delta/2)$ and mean 1, X_2 has support $(0, \delta)$ and mean $\delta/2$. Then stage 1 consists of a single merge operation resulting in a variable with support $(0, 1 + 3\delta/2)$ and mean $1 + \delta/2$.)

To prove the two alternatives, we first introduce some notation. For a random variable X_i, let μ_i denote its mean, $(0, v_i)$ its support, and let $s_i = v_i - \mu_i$ denote its *surplus*. Let $s = \sum_{i=1}^{n'} s_i$ denote the total surplus.

Proposition 3 *If the total surplus satisfies $s < \delta$, then $Pr[X \geq \mu + \delta] = 0$.*

Proof: X is maximized when all X_i come up equal to their respective v_i. In this case

$$X = \sum_{i=1}^{n'} (\mu_i + s_i) = \mu + s < \mu + \delta.$$

□

Hence $s < \delta$ implies that the first alternative holds. For the rest of the proof we assume that $s \geq \delta$, and prove that the second alternative holds, namely, $Pr[X < \mu] \geq \min[\delta/(3/2 + \delta), 1/13]$. Here, we consider several cases (that may be further broken into subcases), depending on the outcome of stage 1.

Lemma 4 *If stage 1 ended with a random variable with mean below $1/2$, then $Pr[X < \mu] \geq \delta/(1/2 + \delta)$.*

Proof: In this case, exactly one random variable remains. Let X_1 be nonconstant random variable left, with mean $\mu_1 < 1/2$ and support $\{0, v_1\}$. By our assumption that $s \geq \delta$ we have that $v_1 \geq \mu_1 + \delta$. Note that the event $X_1 = 0$ implies $X < \mu$. Now $Pr[X_1 = 0] = \delta/(\mu_1 + \delta) \geq \delta/(1/2 + \delta)$, because $\mu_1 < 1/2$. □

Hence we may assume that stage 1 ended with all random variables having mean at least $1/2$. The following property will be used in this case.

Proposition 5 *If stage 1 ended with all random variables having mean at least* $1/2$*, then* $\mu_1/2 \leq \mu_{n'} \leq \mu_1 < 3/2$.

Proof: Recall that the random variables are assumed to be ordered with μ_1 being the largest mean and $\mu_{n'}$ being the smallest mean.

If no random variable has mean greater than 1, then we are done. Hence consider the first time that a random variable with mean greater than 1 is created. This happens by merging two random variables, say X_i and X_j. Let $\mu_i \geq \mu_j$ be their means before the merge. By the definition of *merge*, no other variable had mean smaller than μ_i. By the stopping rule for stage 1, $\mu_j < 1/2$. To get a variable with mean greater than 1, we must have $\mu_i > 1/2$. Note that stage 1 ends after the merge, because no variable with mean below 1/2 is left. Hence the new variable created becomes X_1 with $1 < \mu_1 < 1 + 1/2 = 3/2$. But as $\mu_1 \leq 2\mu_i$ and $\mu_{n'} \geq \mu_i$, it follows that $\mu_{n'} \geq \mu_1/2$. □

The following proposition will be used in the proofs of several of our lemmas. It is most effective when $s < \mu_n$, and μ_n is not much smaller than μ.

Proposition 6 *For independent random variables* $X_1, \ldots, X_n$ *with means* $\mu_1 \geq \ldots \geq \mu_n$ *and supports* $\{0, \mu_1 + s_1\}, \ldots, \{0, \mu_n + s_n\}$*, and with* $X = \sum_{i=1}^n X_i$*,* $\mu = \sum_{i=1}^n \mu_i$ *and* $s = \sum_{i=1}^n s_i$*, the following must hold:*

$$Pr[X < \mu - \mu_n + s] \geq \frac{s}{\mu_1 + s}$$

Proof: It suffices for one random variable come up zero to imply $X < \mu + s - \mu_n$. (The inequality is strict becuase only a variable with $s_i > 0$ may come up 0.) Hence:

$$Pr[X \geq \mu + s - \mu_n] = \Pi_{i=1}^n \frac{\mu_i}{\mu_i + s_i} \leq \Pi_{i=1}^n \frac{\mu_1}{\mu_1 + s_i}.$$

Given that $\sum_{i=1}^{n'} = s$ and that $s_i \geq 0$, the above product is maximized when $s_1 = s$ and $s_i = 0$ for all $i > 1$, giving $\mu_1/(\mu_1 + s)$. Hence $Pr[X < \mu - \mu_n + s] \geq s/(\mu_1 + s)$. □

The following lemma illustrates the desired outcome of stage 1.

Lemma 7 *If stage 1 ended with all random variables having mean at least* $1/2$*, and if* $s < \mu_{n'}$*, then* $Pr[X < \mu] \geq \delta/(3/2 + \delta)$*. If in addition* $Pr[X < \mu] < \delta/(1 + \delta)$ *and* $\delta \leq 1/12$ *then one merge operation before the end of stage 1 it must have been the case that* $Pr[X < \mu + \delta] \geq \delta/(1 + \delta)$.

Proof: The surplus s is smaller than the mean of any of the random variables. Using Proposition 6 we then have $Pr[X < \mu] = \frac{s}{\mu_1 + s}$. Using the assumption that $s \geq \delta$ and the fact that $\mu_1 \leq 3/2$ (Proposition 5), we have that $Pr[X < \mu] \geq \delta/(3/2 + \delta)$.

To prove the second part of the lemma, note that if it happens that $\mu_1 \leq 1$ then we have $Pr[X < \mu] \geq \delta/(1 + \delta)$. Hence we may assume that $\mu_1 > 1$, implying in particular that X_1 is the result of the last *merge* operation (see proof of Proposition 5). Let $s' = s - s_1$ be the surplus of all variables except for X_1. Then analysis as above implies that

$$Pr[X \geq \mu] \leq \frac{\mu_2}{\mu_2 + s'} \leq \frac{1}{1+s'}.$$

Hence if $s' \geq \delta$, $Pr[X < \mu] \geq \delta/(1+\delta)$. So we can assume that $s' < \delta$.

Let us backtrack the last merge operation. Hence instead of X_1 we have two variables X_i and X_j that were merged to give X_1. Let their means be $\mu_i \geq \mu_j$, and their surpluses be s_i and s_j. Observe that necessarily $\mu_j < 1/2$ (otherwise the merge operation would not have been performed), and then the assumption that $\mu_1 > 1$ implies that $\mu_i > 1/2$. As the total surplus of all random variables except for X_i and X_j is $s' < \delta$, we must have $X_i + X_j$ come up larger than $\mu_i + \mu_j$ for $X \geq \mu + \delta$. We consider now two cases.

Case 1: $s_i > 2\delta$**.** Then $Pr[X_i = 0] = \frac{s_i}{\mu_i + s_i} \geq \frac{2\delta}{\mu_i + 2\delta}$. If $X_i = 0$, then in order to have $X_i + X_j > \mu_i + \mu_j$ we must have $X_j > \mu_i + \mu_j$. But this happens with probability at most $\frac{\mu_j}{\mu_i + \mu_j} \leq \frac{1/2}{\mu_i + 1/2}$. Hence

$$Pr[X < \mu + \delta] \geq \frac{2\delta}{\mu_i + 2\delta} \cdot \frac{\mu_i}{\mu_i + 1/2} \geq \frac{\delta}{1+\delta}$$

where the last inequality holds for $\delta \leq 1/2$ because $1/2 \leq \mu_i \leq 1$.

Case 2: $s_i \leq 2\delta$**.** Define $s'' = s' + s_i$ as the surplus of all random variables except for s_j, and observe that $s'' < 3\delta \leq 1/4$, the last inequality holding for $\delta \leq 1/12$.

- If $s'' < \delta$ and $s_j < 1/2$ then it suffices for one random variable to come up 0 to ensure $X < \mu + \delta$. As necessarily $s'' + s_j \geq \delta$ and $\mu_k \leq 1$ for all k, this happens with probability at least $\delta/(1+\delta)$.

- If $s'' < \delta$ and $s_j \geq 1/2$ then it suffices for X_j to come up 0 to ensure $X < \mu + \delta$. This happens with probability at least $1/2$.

- If $\delta \leq s'' < \delta + \mu_j$ and $s_j \geq \delta$ then it suffices for X_j to come up 0 to ensure $X < \mu + \delta$. This happens with probability at least $\delta/(1/2 + \delta)$.

- If $\delta \leq s'' < \delta + \mu_j$ and $s_j < \delta$ then it suffices for some random variable other than X_j to come up 0 to ensure $X < \mu + \delta$. (Recall also that $s'' \leq 1/4$.) This happens with probability at least $\delta/(1+\delta)$.

- If $s'' \geq \delta + \mu_j$ then there is probability of at least $\frac{\delta + \mu_j}{1 + \delta + \mu_j}$ for a random variable other than X_j to come up 0. Thereafter X_j must come up at least $\mu_j + \delta + 1/2 - s'' \geq \mu_j + \delta + 1/4$ for $X \geq \mu + \delta$. The probability of this is at most $\mu_j/(\mu_j + 1/4 + \delta)$. Hence

$$Pr[X < \mu + \delta] \geq \frac{\delta + \mu_j}{1 + \delta + \mu_j} \cdot \frac{1/4 + \delta}{\mu_j + 1/4 + \delta} \geq \frac{\delta}{1+\delta}$$

where the last inequality holds when $4\delta^2 + 4\mu_j\delta \leq 1$, which is true for our parameters of $\mu_j \leq 1/2$ and $\delta \leq 1/12$.

□

Summarizing, the situation so far is that we may assume that $\mu_1/2 \leq \mu_{n'} \leq \mu_1$ $s \geq \mu_{n'}$. We shall prove that in this case $Pr[X < \mu] \geq 1/13$. To prove this, we perform stage 2 of our sequence of transformations. It is composed of a modified form of the merge operations, that we call *modified merge*. The modification will allow us to deal with the event $X < \mu$ rather than $X < \mu + \delta$. Recall that the *reduce support* operation was based on a linear program that minimized $Pr[X < \mu + \delta]$ (via the definition of the q_j). Modify the *reduce support* operation by modifying the objective function of the linear program to be $Pr[X < \mu]$ (by making the respective change in the definition of q_j). Use this *modified reduce support* rather than the original *reduce support* as the second step of *modified merge*. Now *modified merge* does not increase $Pr[X < \mu]$.

Note that an application of *modified merge* may result in a random variable whose mean is smaller than $\mu_1/2$. (For simplicity of notation, we assume that after every step the variables are renamed so as to keep μ_1 the largest mean.) However, even with repeated applications of *modified merge*, there will be at most one such random variable. Let us define $s' = \sum s_i$, where the sum is taken over all random variables whose mean is at least $\mu_1/2$. In particular, at the time when stage 1 ends, $s = s'$.

Stage 2. Apply *modified merge* (on the two random variables with currently lowest mean) until either it cannot be applied anymore (because the number on nonconstant random variables was reduced to one), or the condition $s' \leq \alpha\mu_1$ has been reached, for some constant $0 < \alpha < 1/2$ that will be determined later. This takes at most $n - 1$ steps, because with each application of *merge*, the number of random variables decreases.

Lemma 8 *When $\alpha = 1/3$, then either at the time stage 2 ends or one* modified merge *operation before stage 2 ends*

$$Pr[X < \mu] \geq 1/13$$

The proof of Lemma 8 involves a detailed case analysis (using ideas similar to those that appear in the proof of Lemma 7) and appears in the appendix.

This completes the proof of Theorem 1. □

3 Some remarks

It is straightforward to modify inequality (1) so that there is no formal requirement that the random variables are nonnegative, or that their mean is bounded by 1. Let w be the maximum over all random variables $X_1, \ldots, X_n$ of the respective $\mu_i - l_i$, where l_i is the lowest value in the support of X_i. Then

$$Pr[X \leq \mu + \delta w] \geq \min[\delta/(1 + \delta), 1/13] \quad (2)$$

The constant $1/13$ in Theorem 1 is not best possible, and can be improved with more detailed case analysis. We suspect that the true constant should be $1/e$. Presumably, the way to prove a tight result is to find a sequence of transformations on the random variables that does not increase $Pr[X < \mu+\delta]$, and that gradually brings them to the conjectured worst case for $[X < \mu+\delta]$. The sequence of transformations performed in our proof of Theorem 1 manages to achieve this only when $\delta \leq 1/12$ (or some other constant not far from $1/12$). However, it fails to characterize the worst case for the perhaps more interesting $\delta = 1$. The idea in the proof is to transform the random variables into a situation where a case analysis becomes manageable, at the possible cost of giving up the tightness of the bound. The main principles used are reducing the support of every random variable to two values, getting all random variables (perhaps except one) to have roughly the same mean, reducing the surplus to be of order of magnitude comparable to this mean, and extracting from arbitrarily many random variables a single event of interest, as done in Proposition 6. It should be clear to the reader that more detailed case analysis would provide tighter results. But let us point out some limitations that relate to Lemma 8. As long as one chooses α not larger than $\mu_1/2$ (and in fact, not larger than $3\mu_1/2$), and analyses only the situation at the end of stage 2 or one step earlier, one cannot obtain a bound better than $Pr[X < \mu] \geq 2/9$. For example, assume that during stage 2 we are left with three variables, each with support $(0, \mu)$ and mean $\mu/3$. At this point, $Pr[X < \mu] = (2/3)^3 = 8/27 < 1/e$. After a merge operation, this probability decreases further to $(1/3)\cdot(2/3) = 2/9$. One merge operation later, stage 2 ends. Hence to get (nearly) tight results using the current approach, one may need to modify the definition of stage 2, and perform much more extensive (possibly computer assisted) case analysis.

Acknowledgements

I thank Johan Hastad, Michael Langberg, Eran Ofek, Gideon Schechtman, Benjy Weiss and Avi Wigderson for their interest and involvement in various stages of this work.

References

[1] L. Dubins and L. Savage. *Inequalities for Stochastic Processes. (How to gamble if you must.)* Dover Publications, New York, 1976.

[2] U. Feige. "Sampling vertex degrees and estimating network load parameters." *Manuscript*, 2003

[3] A. Siegel. "Median bounds and their application". *Journal of Algorithms*, 38(1), 184–236, 2001.

A Appendix

Here we prove Lemma 8.

Lemma 9 *If stage 2 ends without the condition $s' \leq \alpha\mu_1$ being reached, then $Pr[X < \mu] > \alpha/(1+\alpha)$.*

Proof: In this case we have only one nonconstant random variable, X_1, with support $\{0, \mu_1 + s'\}$.

$$Pr[X < \mu] = Pr[X_1 = 0] = \frac{s'}{\mu_1 + s'} > \frac{\alpha}{1+\alpha}.$$

$\square$

Lemma 10 *If stage 2 ends with $\beta\mu_1 \leq s' \leq \alpha\mu_1$, where $0 \leq \beta \leq \alpha$ is some constant that will be optimized later, then*

$$Pr[X < \mu] \geq \min\Big[\Big(\frac{\alpha - 2\alpha^2}{1+\alpha}\Big)\ ,\ \Big(\frac{\beta - 2\beta^2}{1+\beta}\Big)\Big].$$

Proof: Consider first only the random variables with mean at least $\mu_1/2$, let X' be their sum and let μ' be the expectation of X'. Over these random variables, the surplus is $s' = \gamma\mu_1$, with $\beta \leq \gamma \leq \alpha < 1/2$. By Proposition 6,

$$Pr[X' < \mu' - (1/2 - \gamma)\mu_1] \geq \frac{s}{\mu_1 + s} = \frac{\gamma}{1+\gamma}.$$

The event $X' < \mu' - (1/2-\gamma)\mu_1$ does not yet imply that $X < \mu$. There still might be one variable $X_{n'}$ with $\mu_{n'} < \mu_1/2$. If $X_{n'}$ turns out $\mu_{n'} + s_{n'}$ and $s_{n'} \geq (1/2-\gamma)\mu_1$ then it still may hold that $X \geq \mu$.

Let us first assume that $\mu_{n'} \leq s' = \gamma\mu_1$. Then by Markov's inequality,

$$Pr[X_{n'} \geq \mu_{n'} + s_{n'}] \leq \frac{\mu_{n'}}{\mu_{n'} + s_{n'}} \leq \frac{\gamma\mu_1}{\gamma\mu_1 + (1/2-\gamma)\mu_1} \leq 2\gamma.$$

Hence

$$Pr[X < \mu] \geq \frac{\gamma}{1+\gamma} \cdot (1 - 2\gamma) = \frac{\gamma - 2\gamma^2}{1+\gamma}.$$

For $0 < \beta \leq \gamma \leq \alpha < 1/2$, the expression above is minimized when $\gamma \in \{\alpha, \beta\}$.

We are left with the case that $\mu_{n'} > s'$. But then we have

$$Pr[X < \mu] \geq Pr[X_{n'} = 0] \geq \frac{s_{n'}}{\mu_{n'} + s_{n'}} \geq \frac{1/2 - \gamma}{1 - \gamma}$$

where we have used the facts that $\mu_{n'} < \mu_1/2$ and $s_{n'} \geq (1/2-\gamma)\mu_1$. As $\gamma \leq \alpha$, we have that $Pr[X < \mu] > (1/2 - \alpha)/(1-\alpha)$. But this probability is larger than $(\alpha - 2\alpha^2)/(1+\alpha)$ of the previous case, and hence can be ignored. $\square$

Lemma 11 *If stage 2 ends with $s' < \beta\mu_1$, and $0 < \beta < \alpha/2$, then one merge prior to the end of stage 2 it must have been the case that $Pr[X < \mu]$ was at least the minimum of the following expressions:*

1. $\frac{\alpha-\beta}{1/2+\alpha-\beta} \cdot \frac{1/2-\beta}{1-\beta}$
2. $\frac{\alpha-3\beta/2}{1+\alpha-3\beta/2} \cdot \frac{1-3\beta/2}{3/2-3\beta/2}$
3. $\frac{\alpha-2\beta}{1+\alpha-2\beta}$
4. $\left(\frac{1/2-\beta}{1-\beta}\right)^2$
5. $\frac{1/2-3\beta/2}{3/2-\beta} \cdot \frac{1-3\beta/2}{3/2-\beta}$

Proof: Consider the last two random variables to have been merged, say X_i and X_j, with means $\mu_i \geq \mu_j$, and let μ_1' be the largest mean at the time. After the *modified merge* of X_i and X_j, the largest mean μ_1 may still have been μ_1', but it could also be as high as $\mu_i + \mu_j$, if this happens to be higher than μ_1'. In fact, μ_1 may also be lower than μ_1', if only one variable is left at the end of stage 2, and this variable underwent an *align with 0* operation. However, in this case the bounds that we get for $X < \mu$ are much stronger than what we get otherwise (details omitted), so we shall ignore this case.

We analyse the situation one merge operation before the end of stage 2. Note that we know that at that time, $s' \geq \alpha\mu_1$, because otherwise stage 2 would have ended earlier. Likewise, the sum $\sum s_i$ taken over all variables except X_i and X_j is at most $\max[\beta\mu_1', \beta(\mu_i + \mu_j)]$, because otherwise we could not have had $s' < \beta\mu_1$ at the end of stage 2. We consider now two cases.

Case 1: $\mu_j < \mu_1'/2$**.** Hence μ_j did not contribute to s'. Note that $\mu_j \geq mu_1'/2$ and follows that $s_i \geq \alpha\mu_1' - \beta\mu_1$. Hence at that point,

$$Pr[X_i = 0] \geq \frac{\alpha\mu_1' - \beta\mu_1}{\mu_i + \alpha\mu_1' - \beta\mu_1}$$

If $X_i = 0$ then in order to have $X \geq \mu$, X_j must contribute at least $\mu_i - \beta\mu_1$ beyond μ_j to X. This may happen with probability at most $\mu_j/(\mu_j + \mu_i - \beta\mu_1)$. We then have

$$Pr[X < \mu] \geq \frac{\alpha\mu_1' - \beta\mu_1}{\mu_i + \alpha\mu_1' - \beta\mu_1} \cdot \frac{\mu_i - \beta\mu_1}{\mu_j + \mu_i - \beta\mu_1}$$

The above expression is minimized when μ_j is maximized (note that increasing μ_j may allow us to increase μ_1, though we are not forced to do so), namely, when $\mu_j = \mu_1'/2$. As $\mu_j \geq \mu_1'/2$, it follows that $\mu_i + \mu_j \geq \mu_1'$. The expression above is minimized when μ_1 is maximized, namely, $\mu_1 = \mu_j + \mu_i]$. Normalising μ_1' to 1, and keeping the notation μ_i to denote μ_i/μ_1', we have after some rearrangements

$$Pr[X < \mu] \geq \frac{\alpha - \beta/2 - \beta\mu_i}{(1-\beta)\mu_i + \alpha - \beta/2} \cdot \frac{(1-\beta)\mu_i - \beta/2}{(1-\beta)\mu_i + 1/2 - \beta/2}$$

The expression above is defined for all $\mu_i \geq 0$. It equals 0 for $\mu_i = \{(\alpha - \beta/2)/\beta, \beta/2(1-\beta)\}$ and positive in between. Moreover, there are only two points where the derivative with respect to μ_i of this expression vanishes (as it is a ratio of two quadratics), and for $\beta < 2\alpha/3$ the expression is ositive in the allowed range of $1/2 \leq \mu \leq 1$. It follows that the expression is minimized when $\mu_i \in \{0, 1, \}$, giving

$$Pr[X < \mu] \geq \min \Big[\frac{\alpha - \beta}{1/2 + \alpha - \beta} \cdot \frac{1/2 - \beta}{1 - \beta}\ ,\ \frac{\alpha - 3\beta/2}{1 + \alpha - 3\beta/2} \cdot \frac{1 - 3\beta/2}{3/2 - 3\beta/2}\Big]$$

This gives items 1 and 2 of the lemma.

Case 2: $\mu_j \geq \mu_1'/2$. Hence both s_i and s_j did contribute to s' (before the last merge), and moreover, $\mu_i + \mu_j \geq \mu_1'$. As in case 1, the worst possibility here is that $\mu_1 = \mu_i + \mu_j$. To simplify notation and without loss of generality we may assume that $\mu_1' = 1$, and then $1/2 \leq \mu_j \leq \mu_i \leq 1$. We have that $s_i + s_j > \alpha - \beta(\mu_i + \mu_j) \geq \alpha - 2\beta$. We are guaranteed than $X < \mu$ if $X_i + X_j < \mu_i + \mu_j - \beta(\mu_i + \mu_j)$. Hence let B denote the event $[X_i + X_j < (1-\beta)(\mu_i + \mu_j)]$. We perform now a subcase analysis for $Pr[B]$.

1. It suffices that either $X_i = 0$ or $X_j = 0$ for B to hold. In this case, using $\mu_i \leq \mu_1$, $s_i + s_j \geq \alpha - 2\beta$, Proposition 6 implies that

$$Pr[B] \geq \frac{\alpha - 2\beta}{1 + \alpha - 2\beta}$$

This gives item 3 in the statement of the lemma.

2. B holds iff $X_i = \mu_i + s_i$. In this subcase necessarily $\mu_i + s_i \geq (1-\beta)(\mu_i + \mu_j)]$. Using the fact that $\mu_i \leq 1$ and $\mu_j \geq 1/2$ we have

$$Pr[B] \geq \frac{1/2 - 3\beta/2}{3/2(1-\beta)}$$

This subcase is dominated by the subcase above and hence can be ignored.

3. B holds iff $X_j = \mu_j + s_j$. This subcase is dominated by the subcase above and can be ignored.

4. B holds unless both $X_i = 0$ and $X_j = 0$. Then necessarily $\mu_i + s_i \geq (1-\beta)(\mu_i + \mu_j)$ and $\mu_j + s_j \geq (1-\beta)(\mu_i + \mu_j)$. We have

$$Pr[B] \geq \Big(\frac{(1-\beta)(\mu_i + \mu_j) - \mu_i}{(1-\beta)(\mu_i + \mu_j)}\Big) \cdot \Big(\frac{(1-\beta)(\mu_i + \mu_j) - \mu_j}{(1-\beta)(\mu_i + \mu_j)}\Big)$$

For fixed $\mu_i + \mu_j$ this expression is minimized when $\mu_i - \mu_j$ is maximized. Hence either $\mu_i = 1$ or $\mu_j = 1/2$. Thereafter, it can be verified that the expression is minimized when the other mean is either maximized or minimized, giving us three possible local minimum points, $\mu_i, \mu_j \in \{1/2, 1\}$, $\mu_j \leq \mu_i$. Two of these give identical values (the cases that $\mu_i = \mu_j$), hence we obtain

$$Pr[B] \geq \min\left[\left(\frac{1/2-\beta}{1-\beta}\right)^2, \left(\frac{1/2-3\beta/2}{3/2-\beta} \cdot \frac{1-3\beta/2}{3/2-\beta}\right)\right]$$

This gives items 4 and 5 in the statement of the lemma.

□

Summing up, we see that after stage 2, $Pr[X < \mu]$ is at least the smallest of the following quantities (where $0 < \beta < \alpha/2 < 1/4$):

- $\frac{\alpha-2\alpha^2}{1+\alpha}$
- $\frac{\beta-2\beta^2}{1+\beta}$
- $\frac{\alpha-\beta}{1/2+\alpha-\beta} \cdot \frac{1/2-\beta}{1-\beta}$
- $\frac{\alpha-3\beta/2}{1+\alpha-3\beta/2} \cdot \frac{1-3\beta/2}{3/2-3\beta/2}$
- $\frac{\alpha-2\beta}{1+\alpha-2\beta}$
- $\left(\frac{1/2-\beta}{1-\beta}\right)^2$
- $\frac{1/2-3\beta/2}{3/2-\beta} \cdot \frac{1-3\beta/2}{3/2-\beta}$

Choosing (suboptimally) $\alpha = 1/3$ and $\beta = 1/8$ gives $Pr[X < \mu] \geq 1/13$ in all cases.

This completes the proof of Lemma 8.