SHARP TAIL BOUNDS BEYOND TWICE THE MEAN

By

Philipp Strack and Jannik M. Westermann

October 2026

COWLES FOUNDATION DISCUSSION PAPER NO. 2567



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

SHARP TAIL BOUNDS BEYOND TWICE THE MEAN

PHILIPP STRACK[†] AND JANNIK M. WESTERMANN[‡]

ABSTRACT. Let $X_1, \dots, X_n$ be independent nonnegative random variables with $\mathbb{E}[X_i] \leq 1$. We prove that, for every $t \geq 2n+1$,

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq 1 - \left(1 - \frac{1}{t}\right)^n.$$

To prove this, we consider a relaxed optimization problem over a set of sequences of ordered, but non-independent random variables. This allows us to reformulate it recursively as a dynamic programming problem. The bound becomes an equality for the binary i.i.d. random variables satisfying $\mathbb{P}[X_i = 0] = 1 - \frac{1}{t}$ and $\mathbb{P}[X_i = t] = \frac{1}{t}$, which remains a maximizer in the relaxed problem.

1. INTRODUCTION

Let $X_1, \dots, X_n$ be independent nonnegative random variables satisfying $\mathbb{E}[X_i] \leq 1$. We study the extremal problem

$$\sup_X \mathbb{P}[X_1 + \cdots + X_n \geq t],$$

where the supremum ranges over all such collections. Markov's inequality gives the universal bound n/t, but this bound does not use independence. A natural independent construction is instead

$$X_i = \begin{cases} t, & \text{with probability } 1/t, \\ 0, & \text{with probability } 1 - 1/t. \end{cases}$$

For this construction the event that the sum reaches t is the event that at least one variable takes its upper value, and hence

$$\mathbb{P}[X_1 + \cdots + X_n \geq t] = 1 - \left(1 - \frac{1}{t}\right)^n.$$

Date: October 8, 2026.

[†]Yale University. Email: philipp.strack@gmail.com. Philipp Strack gratefully acknowledges financial support by the Cowles Foundation.

[‡]Yale University. Email: jmwesterm@gmail.com.

Our main result shows that this lower bound on the extremal probability is sharp when $t \geq 2n + 1$.

This question is a special case of a classical conjecture of Samuels on sums of independent nonnegative random variables with prescribed means [Sam66]. To describe the conjectured extremizers in the equal-mean case, suppose $t > n$ and fix $k \in \{1, \ldots, n\}$. Let $n-k$ variables be identically equal to 1, and let each of the remaining k variables independently take the value $t - n + k$ with probability $1/(t - n + k)$ and the value 0 otherwise. This gives the tail probability

$$H_k(t) := 1 - \left(1 - \frac{1}{t - n + k}\right)^k.$$

The equal-mean specialization of Samuels' conjecture asserts that the maximal tail probability is $\max_{1 \leq k \leq n} H_k(t)$. The construction above corresponds to $k = n$.

Markov's inequality settles the case $n = 1$ and $n = 2$ was solved by Hoeffding and Shrikhande [HS55]. Samuels proved his conjecture for $n = 3, 4$ [Sam66, Sam68], and, when $n \geq 5$, in the far-tail range $t \geq n(n - 1)$ [Sam69].

For *identically* distributed summands, it can be shown that the $k = n$ construction is optimal for $t \geq \frac{5n-2}{3}$, see Frankl and Kupavskii [FK22, Corollary 26] who use connections to fractional matching in hypergraphs. These were established by Alon et. al. [AFH$^+$12] and used by Łuczak, Mieczkowska and Šileikis [ŁMŠ17] to establish the result for $t \geq 2n - 1$.

Our theorem treats arbitrary independent, *non-identically distributed* summands and lowers the previously known quadratic sufficient threshold to a linear one, like there is in the i.i.d. case.

The main idea of our proof is to enlarge the class of admissible random variables until the optimization becomes recursive, while retaining enough of the structure of independence to obtain a sharp bound. We first reduce to two-point distributions, shift their lower support points to zero, and order their upper support points decreasingly. Only then do we remove independence. The resulting adapted processes have a predictable mean budget and conditionally binary jumps whose possible sizes decrease over time. This last monotonicity condition is essential: without it, the relaxed problem recovers the Markov bound. The relaxed problem admits a dynamic programming formulation, which we control by combining a binomial concentration bound for small jumps with an explicit supersolution for large jumps.

Among further work published related to Samuels' conjecture there was interest in the special case of $t = n + 1$, started by Feige [Fei06] after whom this special case of Samuels'

conjecture is named and subsequent improvements by He, Zhang and Zhang [HZZ10] and Garnett [Gar20].

1.1. **Concurrent Work.** Recently, independent of our analysis, Ling [Lin26] has resolved Samuels' conjecture. After the reduction to a sum of binary distributions T and the threshold normalized to 1, the value $\mathbb{E}[T \mid T < 1]$ is exploited. A recursion is set up via

$$\mathbb{P}[T < 1] = \mathbb{P}[T_{-i} < 1]\,(1 - \mathbb{P}[T \geq 1 \mid T_{-i} < 1])$$

where T_{-i} is the sum of the $n-1$ variables excluding i. The value $\mathbb{P}[T_{-i} < 1]$ is bounded recursively and the value $\mathbb{P}[T \geq 1 \mid T_{-i} < 1]$ is controlled using the difference $\mathbb{E}[T \mid T < 1] - \mathbb{E}[T_{-i} \mid T_{-i} < 1]$. This difference is positive for the correct choice of i. The recursion allows bounding $\mathbb{P}[T < 1]$ by a deterministic function depending on $\mathbb{E}[T \mid T < 1]$ over which optimization occurs for the final bound. Han [Han26] provides an alternative presentation of Ling's proof.

Furthermore, independent of our analysis, Fu et al. [FHW$^+$26] have recently also shown that for $t \geq n+1$, it holds $\mathbb{P}[\sum_{i=1}^n X_i \geq t] \leq 1 - (1 - \frac{1}{t})^n$ for $X_1, \ldots, X_n$ independent nonnegative and mean at most one random variables. The proof in Fu et al. is based on a conjecture of Gaffke [Gaf05], recently proved by Vlassis and Thomas [VT26]. The result is: Let $K(x) = \mathbb{P}[\sum_{i=1}^n (x_i - 1)E_i \leq E_0]$ where $E_0, \ldots, E_n \sim^{i.i.d.} \mathrm{Exp}_1$ are exponential 1 distributed. Then for $X_1, \ldots, X_n$ independent, nonnegative with mean at most one and every $\alpha \in [0,1]$ it holds

$$\mathbb{P}[K(X) \leq \alpha] \leq \alpha.$$

Now K can be interpreted as a partial volume of the standard n-simplex. Fu et al. combine Gaffke's conjecture with a generalization by Letwin and Yaskin [LY24] of Grünbaum's inequality [Grü60] for volumes of intersections of half-spaces with convex bodies. This gives the desired formula.

Also based on these results, there appeared a proof of Feige's conjecture, i.e. the special case $t = n+1$, by Nie and Wei [NW26].

1.2. **Remark on Hypothesis testing.** Consider the composite null hypothesis that $X_1, \ldots, X_n$ are independent and nonnegative and satisfy $\mathbb{E}[X_i] \leq 1$ for every i. For $0 < \alpha \leq 1/3$, the test that rejects when $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i \geq \bar{t}_{\alpha,n}$ where the critical threshold is given by

$$\bar{t}_{\alpha,n} := \frac{1}{n(1 - (1-\alpha)^{1/n})},$$

	n										
Method	1	2	3	4	5	6	7	8	9	10	∞
$\alpha = 0.05$											
Sharp	20.000	19.747	19.663	19.621	19.596	19.579	19.567	19.558	19.551	19.546	19.496
Markov	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000	20.000
$\alpha = 0.10$											
Sharp	10.000	9.743	9.659	9.617	9.592	9.575	9.563	9.554	9.547	9.541	9.491
Markov	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000

TABLE 1. Critical values $\bar{t}_{\alpha,n}$ for tests that reject when $\bar{X}_n \geq \bar{t}_{\alpha,n}$.

has size of at most α (i.e. probability of falsely rejecting of at most α). This follows as $\alpha \leq 1/3$ implies $n\bar{t}_{\alpha,n} \geq 2n+1$.[1] Without assuming that $X_1, \ldots, X_n$ are independent, the best possible threshold one can achieve is given by Markov's inequality and thus equal to[2]

$$\bar{t}^{Markov}_{\alpha,n} := \frac{1}{\alpha}\,.$$

Table 1 provides a comparison of the threshold applied to the mean for small samples for tests at the 5% and 10% level. As one can see, assuming that the draws are independent, or even assuming i.i.d. leads only to a small improvement over not making any assumption on the correlation structure. Taylor expanding $\bar{t}_{\alpha,n}$ yields that

$$\bar{t}_{\alpha,n} = \frac{1}{|\log(1-\alpha)|} + \frac{1}{2n} + O(n^{-2}), \qquad \frac{\bar{t}_{\alpha,n}}{\bar{t}^{Markov}_{\alpha,n}} \longrightarrow \frac{\alpha}{|\log(1-\alpha)|}.$$

Moreover, as $n \mapsto \bar{t}_{\alpha,n}$ is decreasing, this bound holds for all n and as $\alpha/|\log(1-\alpha)| = 1 - \alpha/2 - \alpha^2/12 + O(\alpha^3)$ this establishes that assuming i.i.d. can for any n only lead to small improvements (of around $\alpha/2$) over the Markov bound.

The intuitive reason why assuming i.i.d. on its own does not lead to major improvement over the Markov bound lies in the fact that the i.i.d. assumption does not keep the random variables $X_1, \ldots, X_n$ from becoming more and more dispersed.

In fact the bound is attained for X such that

$$\mathbb{E}[|X_1|^k] = \frac{1}{(1-(1-\alpha)^{1/n})^{k-1}} = \left(\frac{n}{|\log(1-\alpha)|}\right)^{k-1} + O(n^{k-2})\,,$$

and thus all moments with $k \geq 2$ of X_1 explode when n grows large.

[1] Bernoulli's inequality gives $(1+1/(2n))^n \geq 3/2$, so $(1-\alpha)^{1/n} \geq (2/3)^{1/n} \geq 2n/(2n+1)$ and hence $n\bar{t}_{\alpha,n} \geq 2n+1$.

[2] To see this denote $b_i = n/\alpha - (n-i)$ and consider X_i which is distributed according to $\frac{1}{b_i}\delta_{b_i} + (1-\frac{1}{b_i})\delta_0$ if $0 = X_1 = X_2 = \ldots = X_{i-1}$ or $i = 1$ and otherwise is equal to 1 with probability 1. These X have the property that $\bar{X}_n \in \{0, 1/\alpha\}$ a.s. which as $\mathbb{E}[\bar{X}_n] = 1$ implies that $\mathbb{P}[\bar{X}_n = \frac{1}{\alpha}] = \alpha$.

2. Main Result and Proof Strategy

Theorem 1. *Let $n \in \mathbb{N}$, and let $X_1, \dots, X_n$ be independent nonnegative random variables satisfying $\mathbb{E}[X_i] \le 1$. Then, for every $t \ge 2n+1$,*

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \ge t\right] \le 1 - \left(1 - \frac{1}{t}\right)^n.$$

Equality is attained by i.i.d. random variables satisfying $\mathbb{P}[X_i = 0] = 1 - 1/t$ and $\mathbb{P}[X_i = t] = 1/t$.

A reduction to independent random variables with binary support. Adding the deterministic quantity $1 - \mathbb{E}[X_i]$ to each summand shows that it suffices to consider variables with mean exactly one.[3] Standard extreme-point arguments then reduce the optimization, without changing its supremum, to two-point distributions; see, for example, [Hoe55, Win88]. Write the support of X_i as $\{a_i, b_i\}$ and set

$$Z_i := X_i - a_i, \qquad m_i := \mathbb{E}[Z_i] = 1 - a_i, \qquad w := \sum_{i=1}^{n} m_i.$$

The variables Z_i have lower support point zero, and

$$\left\{\sum_{i=1}^{n} X_i \ge t\right\} = \left\{\sum_{i=1}^{n} Z_i \ge t - n + w\right\}.$$

Thus shifting the support creates a trade-off: a smaller threshold is purchased at the cost of a smaller total mean budget w. We have thus transformed the problem into an equivalent one in which it is without loss to consider only binary random variables Z_i with support $\{0, u_i\}$. Since the Z_i are independent, we may reorder them so that their upper support points u_i are decreasing.

Relaxing independence. We next relax the problem by considering non-independent random variables. Let I denote the class of sequences of independent random variables and for $u \in (0, \infty]$ and $w \ge 0$, let $M(u, w)$ be the class of nonnegative random processes $Z = (Z_1, Z_2, \dots)$ for which there are nonnegative random variables (U_i) which are predictable in the natural filtration of Z such that

(i) the conditional law of Z_i given $Z^{i-1} := (Z_1, \dots, Z_{i-1})$ is supported on $\{0, U_i\}$;

[3]Indeed, setting $\widetilde{X}_i := X_i + 1 - \mathbb{E}[X_i]$ preserves independence and non-negativity, gives $\mathbb{E}\left[\widetilde{X}_i\right] = 1$, and yields $\widetilde{X}_i \ge X_i$ pointwise. Hence $\mathbb{P}[\sum_{i=1}^n X_i \ge t] \le \mathbb{P}\left[\sum_{i=1}^n \widetilde{X}_i \ge t\right]$.

(ii) $U_1 \le u$ and $U_{i+1} \le U_i$ almost surely;
(iii) the predictable mean satisfies $\mathbb{E}[Z_i \mid Z^{i-1}] \le 1$ and $\sum_{i=1}^{\infty} \mathbb{E}[Z_i \mid Z^{i-1}] \le w$ almost surely.

The variable u bounds the next possible jump, while w is the remaining budget for future means. Let $D_{\le 1}$ be the set of independent, nonnegative sequences of random variables with mean at most 1.

Proposition 1 (Relaxation)**.** *For $n \in \mathbb{N}_0$ and $t \in \mathbb{R}$,*

$$\begin{aligned}\sup_{(X_i)\in D_{\le 1}} \mathbb{P}\left[\sum_{i=1}^{n} X_i \ge t\right] &= \sup_{w\in[0,n]} \sup_{(Z_i)\in I\cap M(\infty,w)} \mathbb{P}\left[\sum_{i=1}^{n} Z_i \ge t-n+w\right] \\ &\le \sup_{w\in[0,n]} \sup_{(Z_i)\in M(\infty,w)} \mathbb{P}\left[\sum_{i=1}^{n} Z_i \ge t-n+w\right].\end{aligned}$$

Proof. The first equality holds following the argument given above, while dropping independence in the second line gives an upper bound. □

Proposition 1 relaxes exactly one constraint: the conditional distribution of each Z_i may now depend on the preceding outcomes. It retains the two properties inherited from the independent problem: the pathwise nesting of the support hulls $[0, U_i]$ and the total predictable mean budget. The results below show that this relaxation is tight in the range $t \ge 2n+1$.

The decreasing-jump condition imposed on the processes in $M(u, w)$ is what makes the removal of independence useful. To see this, suppose it were omitted and $t > n$. As long as all previous outcomes are zero, an adapted process could let $Z_i = t-n+i$ with probability $1/(t-n+i)$ and let $Z_i = 0$ otherwise; after the first non-zero outcome it could set all remaining variables equal to one. Every conditional mean is one, and the sum reaches t if and only if some non-zero outcome occurs. Consequently its tail probability is

$$1 - \prod_{i=1}^{n}\left(1 - \frac{1}{t-n+i}\right) = \frac{n}{t},$$

the Markov bound. Along the all-zero history, however, its possible jump sizes increase. The monotonicity inherited from reordering the independent variables rules out precisely this behavior.

A dynamic programming formulation. To compute the upper bound obtained in Proposition 1, define the value function

$$V_n(x,u,w) := \sup_{(Z_i)\in M(u,w)} \mathbb{P}\left[\sum_{i=1}^{n} Z_i \geq x\right]$$

for $n \in \mathbb{N}_0$. Conditioning on the first binary outcome gives the following Bellman recursion.

Lemma 1 (Dynamic programming principle)**.** *The value function V is the unique solution to the recursive equation*

$$V_n(x,u,w) = \sup_{\substack{0<z\leq u \\ 0\leq m\leq \min\{1,w,z\}}} \frac{m}{z} V_{n-1}(x-z,z,w-m) + \left(1-\frac{m}{z}\right) V_{n-1}(x,z,w-m)$$

with initial value $V_0(x,u,w) = \mathbb{1}_{\{x\leq 0\}}$.

Lemma 1 is the standard dynamic programming principle pioneered by Bellman (see for example Proposition 1(a), and Corollary 5.2(a) in [SB78]). We provide a proof in Section 3 to keep our paper self-contained. The recursive formulation of Lemma 1 is simpler than the original optimization problem. Once the continuation value V_{n-1} is known, the step from V_{n-1} to V_n requires only an optimization over the first binary random variable: its support is $\{0,z\}$ with $0 < z \leq u$, and its mean m is at most $\min\{1,w\}$ (and necessarily at most z). The two continuation values are then already encoded by V_{n-1}.

The continuous-budget analogue of the candidate tail probability is

$$G_w(x) := 1 - \left(1-\frac{1}{x}\right)^w$$

for $x \geq 1$. The following proposition which is established in Section 3 gives a uniform estimate on the value function V.

Proposition 2. *Let $w \geq 0$ and $x \geq 2w+1$ and $x > 1$. Then for $n \in \mathbb{N}_0$ we have*

$$V_n(x,\infty,w) = V_n(x,x,w) \leq G_w(x).$$

Jumps larger than the target threshold x can be truncated at x without decreasing the tail probability, which explains the equality in Proposition 2.

Proof of Theorem 1. Given Proposition 2 we can establish Theorem 1 by the chain

$$\begin{aligned}\sup_{(X_i)\in D_{\le 1}} \mathbb{P}\left[\sum_{i=1}^{n} X_i \ge t\right] &\le \sup_{0\le w\le n} V_n(t-n+w,\infty,w)\\ &\le \sup_{0\le w\le n} G_w(t-n+w)\\ &\le G_n(t).\end{aligned}$$

Here, the first inequality follows from Proposition 1 and the definition of the value function. The second inequality is implied by Proposition 2, as if $t \ge 2n+1$ and $w \le n$, then $t-n+w \ge 2w+1$, so Proposition 2 applies. The last inequality follows as $w \mapsto G_w(t-n+w)$ is increasing on $[0,n]$, so its maximum is attained at $w=n$, where it equals $G_n(t)$.[4] □

Thus, to establish Theorem 1 it remains to establish Proposition 2. The Bellman recursion requires bounds on $V_n(x,u,w)$ for all $u \le x$, not only at $u=x$. We divide the possible jump bounds at

$$u_0 := \frac{x+1}{2}.$$

Small-jump regime. For $n \ge 1$ and $u \le u_0$, monotonicity in the jump bound gives

$$V_n(x,u,w) \le V_n(x,u_0,w) \le Q_{n,w}(x),$$

where the last inequality is established in Proposition 3 with

$$Q_{n,w}(x) := \frac{x+1}{x-1}\left(\frac{2w}{x+1} - 1 + \left(1 - \frac{2w}{n(x+1)}\right)^n\right).$$

The underlying concentration argument compares a normalized sum with the sum over a Bernoulli process in convex order and tests it against $(y-1)_+$. This function counts the number of successes beyond the first and is therefore positive only when at least two successes occur. The method goes back to Bentkus [Ben04]; related binomial comparisons for martingales were developed by Pinelis, see e.g. [Pin06]. Proposition 3 adapts these ideas to a pathwise predictable mean budget.

Large-jump regime. For $u_0 \le u < x$, the relevant terms arise from repeated attempts to make a jump of size u. After the first success, only the residual threshold $x-u$ remains, and

[4] To see this recall that $G_s(x) := 1-(1-1/x)^s$. Set $a := t-n \ge 1$ and $F(s) := G_s(a+s)$. Then

$$\frac{\mathrm{d}}{\mathrm{d}s}\log(1-F(s)) = \log\left(1-\frac{1}{a+s}\right) + \frac{s}{(a+s)(a+s-1)} \le \log\left(1-\frac{1}{a+s}\right) + \frac{1}{a+s} \le 0,$$

because $s \le a+s-1$ and $\log(1-r) \le -r$. Hence F is increasing, and $F(w) \le F(n) = G_n(t)$. The endpoints follow by continuity.

its probability can be bounded by Markov's inequality using the remaining mean budget. For integer budget, the time of the first success has geometric weights. Interpolating these weights continuously in the mean budget motivates the function

$$P(x,u,w) := \int_0^w \lambda p^{w-\ell} \min\left\{1, \frac{\ell}{x-u}\right\} \mathrm{d}\ell,$$

where $p = 1-1/u$ and $\lambda = -\log p$. Equivalently, $\lambda p^{w-\ell}\,\mathrm{d}\ell$ is the continuous interpolation of the first-success probability at the point when ℓ units of budget remain, and $\min\{1, \ell/(x-u)\}$ is the Markov bound for reaching the residual threshold.

Four properties make this heuristic bound into a supersolution of the Bellman recursion. First, the small-jump estimate fits below it at the interface:

$$Q_{n,w}(x) \le P(x,u_0,w).$$

Second, $P(x,u,w)$ is increasing in u on $[u_0,x)$. Third, writing $r = w-m$, the key induction inequality is

$$\frac{m}{z}\min\left\{1, \frac{r}{x-z}\right\} + \left(1-\frac{m}{z}\right) P(x,z,r) \le P(x,z,w).$$

It controls the success branch of the Bellman recursion by Markov's inequality and the failure branch by induction. Finally,

$$\sup_{u_0 \le u < x} P(x,u,w) = G_w(x).$$

The endpoint $u = x$ is handled by the corresponding one-step inequality for G. These facts yield Proposition 2, uniformly in the number of remaining periods. The proof below follows this order: relaxation and dynamic programming, the small-jump concentration inequality, comparison of Q, P, and G, and finally the Bellman induction.

3. Proofs of the Remaining Results

Proposition 1 has been proven in the text.

3.1. **Dynamic programming.**

Proof of Lemma 1. The uniqueness follows inductively immediately. The initial condition holds since the empty sum is 0. Assume the result holds for $n-1$. Now let $(Z_i) \in M(u,w)$. Denote by $z > 0$ the upper value of the support of binary variable Z_1 and by $m = \mathbb{E}[Z_1]$ its mean, such that

$$\mathbb{P}[Z_1 = z] = \frac{m}{z}, \quad \mathbb{P}[Z_1 = 0] = 1 - \frac{m}{z}$$

and $0 < z \le u$ and $0 \le m \le \min\{1, w, z\}$. Since the first mean is m, conditionally on either value of Z_1, the future process $(Z_i)_{i\ge 2}$ has predictable mean budget at most $w - m$. Since the upper values in the predictable supports are decreasing, all upper support points can be at most z. Thus for $y \in \{0, z\}$

$$\mathbb{P}\left[\sum_{i=2}^{n} Z_i \ge x - y \,\middle|\, Z_1 = y\right] \le V_{n-1}(x - y, z, w - m).$$

Therefore conditioning on the outcomes of Z_1 we get

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^{n} Z_i \ge x\right] &= \left(1 - \frac{m}{z}\right)\mathbb{P}\left[\sum_{i=2}^{n} Z_i \ge x \,\middle|\, Z_1 = 0\right] + \frac{m}{z}\mathbb{P}\left[\sum_{i=2}^{n} Z_i \ge x - z \,\middle|\, Z_1 = z\right] \\ &\le \left(1 - \frac{m}{z}\right) V_{n-1}(x, z, w - m) + \frac{m}{z} V_{n-1}(x - z, z, w - m).\end{aligned}$$

Since Z_i was chosen arbitrarily, this proves the upper bound.

Conversely, fix z and m and choose ε-suboptimal processes for the two values $V_{n-1}(x, z, w - m)$ and $V_{n-1}(x - z, z, w - m)$. Let Z_1 be z with probability $\frac{m}{z}$ and 0 otherwise. Let the two processes be the continuations conditionally on the corresponding values of Z_1. The resulting process is in $M(u, w)$ and achieves the right hand side up to ε. Taking $\varepsilon \to 0$ proves the reverse inequality. □

3.2. **Concentration inequality.**

Proposition 3. *Let (Y_i) be a stochastic process taking values in $[0, 1]$. Assume that its predictable mean budget is at most $a \in [0, 1]$, that is, $\sum_{i=1}^{\infty} \mathbb{E}\left[Y_i \,|\, Y^{i-1}\right] \le a$. Then, for every $n \in \mathbb{N}$, $y > 1$*

$$\mathbb{P}\left[\sum_{i=1}^{n} Y_i \ge y\right] \le \frac{1}{y - 1}\left(a - 1 + \left(1 - \frac{a}{n}\right)^n\right).$$

In particular, if $u \le u_0 := \frac{x+1}{2}$, $w \ge 0$ and $x \ge 2w - 1$, $x > 1$ then

$$V_n(x, u, w) \le \frac{x + 1}{x - 1}\left(\frac{2w}{x + 1} - 1 + \left(1 - \frac{2w}{n(x + 1)}\right)^n\right) = Q_{n,w}(x).$$

Proof of Proposition 3. Let $\xi_1, \ldots, \xi_n$ be i.i.d. uniformly distributed on the unit interval and independent of the sequence (Y_i). Let

$$B_i := \mathbb{1}_{\{\xi_i \le Y_i\}}, \quad B = \sum_{i=1}^{n} B_i, \quad S = \sum_{i=1}^{n} Y_i.$$

Notice that

$$\mathbb{P}\left[\xi_i \le Y_i \,|\, Y_i\right] = Y_i.$$

Thus $\mathbb{E}\left[B \,|\, Y^n\right] = S$ and hence $\mathbb{E}\left[B \,|\, S\right] = S$, i.e. B is a mean-preserving spread of S. In other words, B dominates S in the convex order. For the convex function $y \mapsto (y-1)_+$ we have by the conditional Jensen's inequality

$$\mathbb{E}\left[(S-1)_+\right] \le \mathbb{E}\left[(B-1)_+\right].$$

Let

$$p_i := \mathbb{P}\left[B_i = 1 \,\middle|\, Y^{i-1}, B^{i-1}\right].$$

Then independence of the ξ_i gives

$$p_i = \mathbb{E}\left[Y_i \,\middle|\, Y^{i-1}\right]$$

and predictable success budget

$$\sum_{i=1}^{n} p_i \le a.$$

In particular, $p_1 = \mathbb{E}\left[Y_1\right]$ is deterministic.

Now set

$$R_n(a) := a - 1 + \left(1 - \frac{a}{n}\right)^n.$$

We prove by induction on n that every such Bernoulli process satisfies

$$\mathbb{E}\left[(B-1)_+\right] \le R_n(a).$$

For $n = 1$, both sides vanish. Suppose the assertion holds for $n-1$ and write

$$B' = \sum_{i=2}^{n} B_i.$$

For almost every value of (Y_1, B_1), under the corresponding conditional law, the remaining process has predictable success budget at most $a - p_1$. Hence

$$\mathbb{E}\left[(B'-1)_+ \,|\, Y_1, B_1\right] \le R_{n-1}(a - p_1) \text{ a.s.}$$

Moreover, the tower property gives

$$\mathbb{E}\left[B' \,|\, Y_1, B_1\right] = \mathbb{E}\left[\sum_{i=2}^{n} p_i \,\middle|\, Y_1, B_1\right] \le a - p_1 \text{ a.s.}$$

Writing $(B-1)_+ = B_1 B' + (1-B_1)(B'-1)_+$, we obtain, using the above inequalities

$$\begin{aligned}\mathbb{E}\left[(B-1)_+\right] &\le p_1(a-p_1) + (1-p_1)R_{n-1}(a-p_1)\\ &= a - 1 + (1-p_1)\left(1 - \frac{a-p_1}{n-1}\right)^{n-1}\\ &\le a - 1 + \left(1-\frac{a}{n}\right)^n = R_n(a)\end{aligned}$$

where the last inequality follows pointwise from AM-GM applied to the n numbers $1-p_1, 1-\frac{a-p_1}{n-1}, \dots, 1-\frac{a-p_1}{n-1}$. This concludes the induction.

Note that as $\{\sum_{i=1}^n Y_i \ge y\} = \{S \ge y\}$ we have

$$(y-1)\mathbb{1}_{\{\sum_{i=1}^n Y_i \ge y\}} \le (S-1)_+.$$

Taking the expectation we get the first part of the proposition

$$(y-1)\,\mathbb{P}\left[\sum_{i=1}^n Y_i \ge y\right] \le \mathbb{E}\left[(S-1)_+\right] \le \mathbb{E}\left[(B-1)_+\right] \le R_n(a).$$

For any admissible process (Z_i) for $V_n(x,u_0,w)$, the result applies to $Y_i = Z_i/u_0$, $a = w/u_0$ and $y = x/u_0$, since $x \ge 2w-1$ guarantees $w \le u_0$, that is $a \le 1$. Thus

$$V_n(x,u_0,w) \le \frac{u_0}{x-u_0} R_n\left(\frac{w}{u_0}\right) = \frac{x+1}{x-1}\left(\frac{2w}{x+1} - 1 + \left(1 - \frac{2w}{n(x+1)}\right)^n\right) = Q_{n,w}(x).$$

Since $V_n(x,u,w) \le V_n(x,u_0,w)$ for $u \le u_0$ we are done. □

3.3. **Order of the bounds $G \ge P \ge Q$.** First two auxiliary lemmas.

Lemma 2. *Let $q \in (0,1)$, $p = 1-q$ and $\lambda = -\log p$. Then*

$$\lambda \ge q.$$

Moreover if $q = \frac{1}{x-s}$ then

$$\lambda'(s) = \frac{q^2}{1-q}, \qquad \frac{\lambda'(s)}{\lambda(s)} \le \frac{q}{1-q} = \frac{1}{x-s-1}.$$

Proof. Taking the logarithm of the inequality $1-q \le e^{-q}$ gives $\lambda \ge q$. If $q = \frac{1}{x-s}$ then $q' = q^2$ and thus $\lambda'(s) = \frac{q'}{1-q} = \frac{q^2}{1-q}$. Dividing by $\lambda \ge q$ gives the remaining claim. □

Lemma 3. *Let*

$$g(y) := \frac{y - 1 + e^{-y}}{y}, \quad y > 0.$$

Then $0 < g(y) < 1$ *and*

$$g'(y) \geq 0, \qquad \frac{g'(y)}{g(y)} \leq \frac{1}{y}, \qquad \frac{g'(y)}{1-g(y)} \leq \frac{1}{2}.$$

Proof. Notice that $g(y) = \int_0^1 (1-e^{-ty})\,\mathrm{d}t \in (0,1)$ and $g'(y) = \int_0^1 te^{-ty}\,\mathrm{d}t \geq 0$. Now it holds $ae^{-a} \leq 1-e^{-a}$ for $a \geq 0$, thus $a = ty$ gives

$$yg'(y) = \int_0^1 tye^{-ty}\,\mathrm{d}t \leq \int_0^1 (1-e^{-ty})\,\mathrm{d}t = g(y).$$

Finally

$$\frac{g'(y)}{1-g(y)} = \frac{\int_0^1 te^{-ty}\,\mathrm{d}t}{\int_0^1 e^{-ty}\,\mathrm{d}t}$$

is an average with decreasing density and therefore positive and less than the average of the uniform distribution on the unit interval which is $\frac{1}{2}$. □

Lemma 4. *For* $w \geq 0$ *and* $x > 1$, $x \geq 2w+1$, $n \in \mathbb{N}$ *we have*

$$Q_{n,w}(x) \leq P(x, u_0, w).$$

Proof. For $w = 0$ the result is immediate so assume $w > 0$. Let us denote $s = x - u_0$, $a = \frac{w}{u_0}$ as well as $q = \frac{1}{u_0} \in (0,1)$, $p = 1-q$ and $\lambda = -\log p$. Since $x \geq 2w+1$ we have $s \geq w$, thus in the definition of P, the Markov bound is active on all of $[0,w]$. That is

$$\begin{aligned} P(x, u_0, w) &= \int_0^w \lambda p^{w-\ell} \min\{1, \frac{\ell}{s}\}\,\mathrm{d}\ell \\ &= \frac{p^w}{s} \int_0^w \lambda p^{-\ell}\ell\,\mathrm{d}\ell \end{aligned}$$

Integration by parts shows the following equality for α, β

$$\int_\alpha^\beta \lambda p^{-\ell}\ell\,\mathrm{d}\ell = p^{-\beta}\left(\beta - \frac{1}{\lambda}\right) - p^{-\alpha}\left(\alpha - \frac{1}{\lambda}\right). \tag{1}$$

Applying it yields since $p = e^{-\lambda}$ and using g from Lemma 3

$$P(x, u_0, w) = \frac{\lambda w - 1 + p^w}{\lambda s} = \frac{w}{s} g(\lambda w).$$

Since generally $\left(1 - \frac{a}{n}\right)^n \leq e^{-a}$, we have

$$R_n(a) = a - 1 + \left(1 - \frac{a}{n}\right)^n \leq a - 1 + e^{-a} = ag(a).$$

Thus

$$Q_{n,w}(x) = \frac{u_0}{s} R_n(a) \leq \frac{u_0}{s} ag(a) = \frac{w}{s} g(a).$$

By Lemma 2 we have $\lambda \geq q = \frac{1}{u_0}$ and thus $\lambda w \geq a$. Since by Lemma 3 g is increasing, we have

$$Q_{n,w}(x) \leq \frac{w}{s} g(a) \leq \frac{w}{s} g(\lambda w) = P(x, u_0, w). \qquad \square$$

Lemma 5. *Let $w \geq 0$ and $x > 1$ and $x \geq 2w + 1$. Then $P(x, u, w)$ is increasing in u on $[u_0, x)$.*

Proof. Write

$$s = x - u,$$

thus $0 < s \leq \frac{x-1}{2}$. We will prove that $P(x, x - s, w)$ is decreasing as a function of s. Write

$$q = \frac{1}{u} = \frac{1}{x - s}, \quad p = 1 - q, \quad \lambda = -\log p$$

and note that $q \in (0, 1)$ since $u \geq u_0 = \frac{x+1}{2}$ and $x > 1$.

Recall that

$$P(x, u, w) = \int_0^w \lambda p^{w-\ell} \min\left\{1, \frac{\ell}{s}\right\} \mathrm{d}\ell.$$

We show that P is decreasing in s in the cases $s \leq w$ and $s \geq w$.

Case 1: $s \leq w$. The minimum equals $\frac{\ell}{s}$ on $[0, s]$ and equals 1 on $[s, w]$. Hence by direct integration for the second term and (1) for the first term we get

$$\begin{aligned} P(x, u, w) &= \frac{p^w}{s} \int_0^s \lambda p^{-\ell} \ell \, \mathrm{d}\ell + p^w \int_s^w \lambda p^{-\ell} \, \mathrm{d}\ell \\ &= 1 - p^{w-s} \frac{1 - p^s}{s\lambda}. \end{aligned}$$

Since $p = e^{-\lambda}$, this means

$$1 - P(x, u, w) = e^{-\lambda(w-s)}(1 - g(s\lambda)).$$

To show that $1 - P$ is increasing in s, we show that its logarithmic derivative is positive. Indeed we have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s} \log(1 - P(x, u, w)) &= -\lambda'(w - s) + \lambda - \frac{g'(s\lambda)}{1 - g(s\lambda)}(\lambda' s + \lambda) \\ &\geq \lambda - \lambda'(w - s) - \frac{1}{2}(\lambda + s\lambda') && \text{Lemma 3} \\ &= \frac{\lambda}{2} - \lambda'(w - \frac{s}{2}) \\ &\geq \frac{q}{2} - \frac{q}{x - s - 1}(w - \frac{s}{2}) && \text{Lemma 2} \\ &\geq 0 && q > 0, \quad x \geq 2w + 1. \end{aligned}$$

Case 2: $s \geq w$. If $w = 0$, then $P(x, u, 0) = 0$ so there is nothing to prove. Assume $w > 0$. Now the minimum equals $\frac{\ell}{s}$ on the whole interval $[0, w]$. Using (1) we get

$$P(x, u, w) = \frac{p^w}{s} \int_0^w \lambda p^{-\ell} \ell \, \mathrm{d}\ell = \frac{w}{s} g(\lambda w).$$

Taking the log derivative we have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s} \log P(x, u, w) &= -\frac{1}{s} + \frac{g'(\lambda w)}{g(\lambda w)} \lambda' w \\ &\leq -\frac{1}{s} + \frac{\lambda'}{\lambda} && \text{Lemma 3} \\ &\leq -\frac{1}{s} + \frac{1}{x - s - 1} && \text{Lemma 2} \\ &\leq 0 && s \leq \frac{x-1}{2}. \end{aligned}$$

Since $P(x, u, w)$ is continuous at $s = w$, this concludes the proof. □

Lemma 6. *Let $w \geq 0$ and $x \geq 2w + 1$ and $x > 1$. Then*

$$\sup_{u \in [u_0, x)} P(x, u, w) = G_w(x).$$

Proof. By Lemma 5 the supremum over $[u_0, x)$ is the limit as $u \to x$. Recall that

$$P(x, u, w) = \int_0^w -\log\left(1 - \frac{1}{u}\right) \left(1 - \frac{1}{u}\right)^{w-\ell} \min\left\{1, \frac{\ell}{x - u}\right\} \mathrm{d}\ell.$$

Now as $u \to x$ we have $1 - \frac{1}{u} \to 1 - \frac{1}{x}$ and $\min\{1, \frac{\ell}{x-u}\} \to 1$ for all $\ell > 0$. Thus dominated convergence gives

$$P(x, u, w) \xrightarrow{u \uparrow x} \int_0^w -\log\left(1 - \frac{1}{x}\right) \left(1 - \frac{1}{x}\right)^{w-\ell} \mathrm{d}\ell = 1 - \left(1 - \frac{1}{x}\right)^w = G_w(x).$$

□

3.4. **Induction.**

Lemma 7. *Let $w \geq 0$ and $0 \leq m \leq \min\{1, w\}$ and denote the remaining budget $r = w - m$. Let $x > 1$ as well as $z \in [u_0, x)$. Then*

$$\frac{m}{z} \min\left\{1, \frac{r}{x - z}\right\} + \left(1 - \frac{m}{z}\right) P(x, z, r) \leq P(x, z, w). \tag{2}$$

Proof. Denote $q = \frac{1}{z} \in (0,1)$, $p = 1-q$ and $\lambda = -\log p$. Using the Bernoulli inequality for an exponent $m \in [0,1]$ we have

$$p^m = (1-q)^m \leq 1 - mq. \tag{3}$$

Note that $h(\ell) = \min\{1, \frac{\ell}{x-z}\}$ is increasing in ℓ. Thus after direct integration

$$P(x,z,r) \leq h(r) \int_0^r \lambda p^{r-\ell} \,\mathrm{d}\ell = h(r)(1-p^r) \leq h(r). \tag{4}$$

Further we bound $P(x,z,w)$ by splitting the integration interval and noting that the first term is $p^{w-r}P(x,z,r) = p^m P(x,z,r)$ and for the second term we bound via h increasing, that is

$$P(x,z,w) = \int_0^r \lambda p^{w-\ell} h(\ell) \,\mathrm{d}\ell + \int_r^w \lambda p^{w-\ell} h(\ell) \,\mathrm{d}\ell \geq p^m P(x,z,r) + (1-p^m)h(r).$$

The term in the second bracket of the following equation is the left term in (2). The inequality

$$[p^m P(x,z,r) + (1-p^m)h(r)] - [(1-mq)P(x,z,r) + mqh(r)] = (1-p^m-mq)(h(r) - P(x,z,r)) \geq 0$$

follows from (3) and (4). Combining the last two inequalities thus completes the proof. □

Lemma 8. *Let $n \in \mathbb{N}_0$ and $w \geq 0$, as well as $x > 1$ and $x \geq 2w+1$ and $u \in [u_0, x)$. Then*

$$V_n(x,u,w) \leq P(x,u,w)$$

Proof. For $n = 0$ we have $V_0(x,u,w) = \mathbb{1}_{\{x \leq 0\}} = 0$. Now for sake of induction assume the lemma holds for $n-1$. We have

$$V_n(x,u,w) = \sup_{\substack{0 < z \leq u \\ 0 \leq m \leq \min\{1,w,z\}}} \frac{m}{z} V_{n-1}(x-z, z, w-m) + \left(1 - \frac{m}{z}\right) V_{n-1}(x,z,w-m)$$

$$= \max\left\{ V_n(x,u_0,w), \sup_{\substack{u_0 < z \leq u \\ 0 \leq m \leq \min\{1,w,z\}}} \frac{m}{z} V_{n-1}(x-z, z, w-m) + \left(1 - \frac{m}{z}\right) V_{n-1}(x,z,w-m) \right\}.$$

By Proposition 3 the first term is bounded by $Q_{n,w}(x) \leq P(x,u,w)$ by Lemma 4 and Lemma 5. For $z \in (u_0, u]$ we get

$$\begin{aligned} &\frac{m}{z} V_{n-1}(x-z,z,w-m) + \left(1 - \frac{m}{z}\right) V_{n-1}(x,z,w-m) \\ &\leq \frac{m}{z} \min\left\{1, \frac{w-m}{x-z}\right\} + \left(1 - \frac{m}{z}\right) P(x,z,w-m) \leq P(x,z,w) \leq P(x,u,w), \end{aligned}$$

where the first inequality uses the Markov inequality and induction hypothesis, the second inequality uses (2) and the last inequality uses Lemma 5. This concludes the proof. □

Proof of Proposition 2. Denote

$$K = \sup_{0\le m\le \min\{1,w\}} \frac{m}{x} + \left(1-\frac{m}{x}\right) V_{n-1}(x,x,w-m).$$

Since $V_n(x,u,w)$ is clearly increasing in u, by Lemma 8 we have

$$V_n(x,x,w) \le \max\{K, \sup_{u\in[u_0,x)} P(x,u,w)\}.$$

Denoting $p = 1-\frac{1}{x} \in (0,1)$ we have, using Bernoulli's inequality for the exponent $m \in [0,1]$,

$$\frac{m}{x} + \left(1-\frac{m}{x}\right) G_{w-m}(x) = 1 - \left(1-\frac{m}{x}\right) p^{w-m} \le 1 - p^m p^{w-m} = G_w(x).$$

We proceed by induction, the case $n=0$ is immediate. By induction hypothesis we have $V_{n-1}(x,x,w-m) \le G_{w-m}(x)$ and thus $K \le G_w(x)$. Lemma 6 gives the inequality.

It remains to show $V_n(x,\infty,w) \le V_n(x,x,w)$. To this end for $Z \in M(\infty,w)$, consider $\tilde{Z}_i = \min\{Z_i, x\}$ and note that $(\tilde{Z}_i) \in M(x,w)$, and since the random variables are nonnegative, $\{\sum_{i=1}^n \tilde{Z}_i \ge x\} = \{\sum_{i=1}^n Z_i \ge x\}$. This concludes the argument. □

References

[AFH+12] N. Alon, P. Frankl, H. Huang, V. Rödl, A. Ruciński, and B. Sudakov, *Large matchings in uniform hypergraphs and the conjectures of Erdős and Samuels*, Journal of Combinatorial Theory, Series A **119** (2012), no. 6, 1200–1215.

[Ben04] V. Bentkus, *On Hoeffding's inequalities*, The Annals of Probability **32** (2004), no. 2, 1650 – 1673.

[Fei06] U. Feige, *On sums of independent random variables with unbounded variance and estimating the average degree in a graph*, SIAM Journal on Computing **35** (2006), no. 4, 964–984.

[FHW+26] W. Fu, Y. Han, G. Wang, J. Yan, P. Zhang, and Z. Zhou, *Sharp small-deviation inequalities for sums of independent nonnegative random variables*, arXiv preprint arXiv:2607.23980, 2026.

[FK22] P. Frankl and A. Kupavskii, *The Erdős matching conjecture and concentration inequalities*, Journal of Combinatorial Theory, Series B **157** (2022), 366–400.

[Gaf05] N. Gaffke, *Three test statistics for a nonparametric one-sided hypothesis on the mean of a nonnegative variable*, Mathematical Methods of Statistics **14** (2005), no. 4, 451–467.

[Gar20] B. Garnett, *Small deviations of sums of independent random variables*, Journal of Combinatorial Theory, Series A **169** (2020), 105119.

[Grü60] B. Grünbaum, *Partitions of mass-distributions and of convex bodies by hyperplanes*, Pacific Journal of Mathematics **10** (1960), no. 4, 1257–1261.

[Han26] Y. Han, *A cap-move reformulation of Ling's proof of Samuels' conjecture*, arXiv preprint arXiv:2608.22816, 2026.

[Hoe55] W. Hoeffding, *The extrema of the expected value of a function of independent random variables*, The Annals of Mathematical Statistics **26** (1955), no. 2, 268–275.

[HS55] W. Hoeffding and S. S. Shrikhande, *Bounds for the distribution function of a sum of independent, identically distributed random variables*, The Annals of Mathematical Statistics **26** (1955), no. 3, 439–449.

[HZZ10] S. He, J. Zhang, and S. Zhang, *Bounding probability of small deviation: A fourth moment approach*, Mathematics of Operations Research **35** (2010), no. 1, 208–232.

[Lin26] Z. Ling, *On Samuels' conjecture*, arXiv preprint arXiv:2608.18392, 2026.

[ŁMŠ17] T. Łuczak, K. Mieczkowska, and M. Šileikis, *On maximal tail probability of sums of nonnegative, independent and identically distributed random variables*, Statistics & Probability Letters **129** (2017), 12–16.

[LY24] B. Letwin and V. Yaskin, *A generalization of Grünbaum's Inequality*, Israel Journal of Mathematics (2024), To appear.

[NW26] Z. Nie and J. Wei, *On Feige's conjecture*, arXiv preprint arXiv:2607.24528, 2026.

[Pin06] I. Pinelis, *Binomial upper bounds on generalized moments and tail probabilities of (super) martingales with differences bounded from above*, Lecture Notes-Monograph Series **51** (2006), 33–52.

[Sam66] S. M. Samuels, *On a Chebyshev-type inequality for sums of independent random variables*, The Annals of Mathematical Statistics **37** (1966), no. 1, 248–259.

[Sam68] ———, *More on a Chebyshev-type inequality for sums of independent random variables*, Tech. Report 155, Department of Statistics, Purdue University, 1968.

[Sam69] ———, *The Markov inequality for sums of independent random variables*, The Annals of Mathematical Statistics **40** (1969), no. 6, 1980–1984.

[SB78] Steven E. Shreve and Dimitri P. Bertsekas, *Alternative theoretical frameworks for finite horizon discrete-time stochastic optimal control*, SIAM Journal on Control and Optimization **16** (1978), no. 6, 953–978.

[VT26] N. Vlassis and P. S. Thomas, *An exact distribution-free test for means of nonnegative random variables*, arXiv preprint arXiv:2607.08415, 2026.

[Win88] G. Winkler, *Extreme points of moment sets*, Mathematics of Operations Research **13** (1988), no. 4, 581–587.