On Sums of Monotone Random Integer Variables

Anders Aamand[*]   Noga Alon[†]   Jakob Bæk Tejs Houen[‡]
Mikkel Thorup[‡]

Abstract

We say that a random integer variable X is *monotone* if the modulus of the characteristic function of X is decreasing on $[0, \pi]$. This is the case for many commonly encountered variables, e.g., Bernoulli, Poisson and geometric random variables. In this note, we provide estimates for the probability that the sum of independent monotone integer variables attains precisely a specific value. We do not assume that the variables are identically distributed. Our estimates are sharp when the specific value is close to the mean, but they are not useful further out in the tail. By combining with the trick of *exponential tilting*, we obtain sharp estimates for the point probabilities in the tail under a slightly stronger assumption on the random integer variables which we call *strong monotonicity*.

1 Introduction

   In this note we provide sharp estimates for the probability that the sum of independent (not necessarily identically distributed) integer-valued random variables attains precisely a specific value. Our estimates hold under a fairly general assumption on the properties of the random variables, which for example is satisfied for Bernoulli, Poisson and geometric random variables. The bounds on the point probabilities derived in this paper have been used to understand the distribution of balls in capacitated bins [1]. In the cleanest combinatorial variant of the problem, where the balls arrive sequentially and each ball picks a uniformly random non-full bin, they just needed the point probabilities of sums of i.i.d. Bernoulli variables. However, for a more dynamic distribution system, they had to apply the bounds for sums of a mix of Bernoulli and geometrically distributed variables.

---

[*]Computer Science and Artificial Intelligence Lab, MIT, Massachusetts, USA. E-mail: aamand@mit.edu. Anders Aamand is funded by DFF-International Postdoc Grant 0164-00022B from the Independent Research Fund Denmark

[†]Department of Mathematics, Princeton University, Princeton, New Jersey, USA and Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv, Israel. E-mail: nalon@math.princeton.edu. Research supported in part by NSF grant DMS-1855464, BSF grant 2018267 and the Simons Foundation.

[‡]Basic Algorithms Research Copenhagen (BARC), University of Copenhagen, Denmark. E-mail: jakn@di.ku. dk,mikkel2thorup@gmail.com. BARC is supported by the VILLUM Foundation grant 16582.

On Sums of Monotone Random Integer Variables

Recall that for a real random variable X, the *characteristic function* of X is the map $f_X : \mathbb{R} \to \mathbb{C}$ given by $f_X(\lambda) = \mathrm{E}[e^{i\lambda X}]$. We say that a real random variable X is *monotone* if $|f_X|$ is decreasing on $[0, \pi]$. In the first part of this note (Section 2), we provide estimates for the point probabilities of a sum, $X = \sum_{j \in [k]} X_j$, of independent monotone random integer variables[1]. To be precise, for any given $t \in \mathbb{Z}$, we estimate the probability $\Pr[X = t]$. Our estimates are sharp whenever t is close to the mean $\mathrm{E}[X]$, but they are not useful further out in the tail. To handle point probabilities in the tail, we require a slightly stronger assumption on the random variables which we now describe.

For a random integer variable X we define $I_X = \{\theta \in \mathbb{R} : \mathrm{E}[e^{\theta X}] < \infty\}$, to consist of those $\theta \in \mathbb{R}$ for which the moment generating function of X is defined. We note that I_X is an interval with $0 \in I_X$. For $\theta \in I_X$, we may define the *exponentially tilted* random variable X_θ by $\Pr[X_\theta = t] = \frac{\Pr[X=t]e^{\theta t}}{\mathrm{E}[e^{\theta X}]}$ for $t \in \mathbb{Z}$. We say that X is *strongly monotone* if (1) $I_X \neq \{0\}$ and (2) X_θ is monotone for each $\theta \in I_X$. In the second part of this note (Section 3), we use the trick of exponential tilting to provide estimates for the point probabilities of a sum of independent strongly monotone random integer variables, $X = \sum_{j \in [k]} X_j$, which are also sharp in the tail.

It follows by direct computation that Bernoulli, Poisson, and geometric random variables are monotone, and moreover, that exponentially tilting these variables again yields Bernoulli, Poisson and geometric variables. In particular, these variables are all strongly monotone, so our results give sharp estimates for the point probabilities of the sum of (a mix of) such variables. In Section 3, we provide examples of the estimates that can be obtained for such a sum using our results.

In the note we will consider the following setting. Let k be an integer and $(X_j)_{j \in [k]}$ independent integer-valued random variables with $\mathrm{E}[X_j] = \mu_j$ and $\mathrm{Var}[X_j] = \sigma_j^2$ for $j \in [k]$. Let $X = \sum_{i \in [k]} X_i$, and further $\mu = \sum_{j \in [k]} \mu_j$ and $\sigma^2 = \sum_{j \in [k]} \sigma_j^2$ be respectively the expectation and variance of X. The main result of the note is the following theorem.

Theorem 1.1. *There exists a universal constant c, such that if X is monotone, then for every t for which $\mu + t\sigma$ is an integer, the probability that X is precisely $\mu + t\sigma$ satisfies,*

$$\left| \Pr[X = \mu + t\sigma] - \frac{1}{\sqrt{2\pi}\sigma}e^{-t^2/2} \right| \leq c \left(\frac{\sum_{j \in [k]} \mathrm{E}\left[|X_j - \mu_j|^3\right]}{\sigma^3} \right)^2. \tag{1.1}$$

Remark 1.2. We note that if each X_j is monotone, then X is as well. Indeed, the characteristic function of X can be factorized as $f_X(\lambda) = \prod_{j \in [k]} f_{X_j}(\lambda)$. In particular, Theorem 1.1 holds when each of the variables $(X_j)_{j \in [k]}$ is monotone.

Our result is reminiscent of the Berry-Esseen theorem, but instead of bounding the distance between the cumulative function of X and the cumulative function of the normal distribution as the Berry-Esseen theorem does, our result bounds the distance between the density function of X and the density function of the normal distribution. This setting has been studied before in the context of large deviation theory, e.g., by Blackwell and Hodges [3] and by Iltis [7] in the d-dimensional case. They do not require X to be monotone but they only consider the case where $(X_j)_{j \in [k]}$ are identically distributed and are interested in the asymptotical behavior when $k \to \infty$. In particular the distribution of the variables $(X_j)_{j \in [k]}$ cannot depend on k. McDonald [8] considers variables that are not necessarily identically distributed but again in the limit $k \to \infty$ and with certain extra assumptions on the distribution of the variables. In this work we are not interested in such asymptotic bounds and our result is a uniform bound for monotone variables.

Another line of related work is the asymptotic expansions in the local limit theorem. See for instance [5] (Section 51) or [6] (Theorems 4.5.3 and 4.5.4) which for any given r

---

[1]We define $[k] = \{0, \dots, k-1\}$

provide expressions for the point probabilities of a sum of k i.i.d. random integer variables $(X_j)_{j\in[k]}$ up to an additive error of $o(k^{-r})$. Again, these results holds asymptotically as $k \to \infty$ and the distributions of the $(X_j)_{j\in[k]}$ cannot depend on k. The case of different distributions is considered in [4] which provide expansions of the point probabilities using regular and trigonometric polynomials up to an additive error of $o(\sigma^{-r})$ (without any assumptions of monotonicity).

2 Point Probabilities Near the Mean

The goal of this section is to prove Theorem 1.1, but before diving into the proof, we provide some examples of random variables for which the condition of the lemma is satisfied. Let $p \in [0, 1]$ and $\lambda > 0$. Let Y be a Bernoulli variable with $\Pr[Y = 1] = 1 - \Pr[Y = 0] = p$, let Z be geometric with $\Pr[Z = k] = p^k(1 - p)$ for $k \in \mathbb{N}_0$, and let W be Poisson with $\Pr[W = k] = \lambda^k e^{-\lambda}/k!$ for $k \in \mathbb{N}_0$. Let f_Y, f_Z, f_W be the characteristic functions for Y, Z, and W. Then for $\lambda \in \mathbb{R}$,

$$f_Y(\lambda) = 1 - p + pe^{i\lambda}, \quad f_Z(\lambda) = \frac{1-p}{1 - pe^{i\lambda}}, \quad \text{and} \quad f_W(\lambda) = e^{\lambda(e^{i\lambda}-1)}.$$

Thus,

$$|f_Y(\lambda)|^2 = (1 - p + pe^{i\lambda})(1 - p + pe^{-i\lambda}) = 1 + 2p(1-p)(\cos\lambda - 1)$$

$$|f_Z(\lambda)|^2 = \left(\frac{1-p}{1 - pe^{i\lambda}}\right)\left(\frac{1-p}{1 - pe^{-i\lambda}}\right) = \frac{(1-p)^2}{1 + p^2 - 2p\cos\lambda}, \quad \text{and}$$

$$|f_W(\lambda)| = e^{\lambda(\cos\lambda - 1)},$$

which are all decreasing functions on $[0, \pi]$.

We will need the following simple Lemma on random integer variables.

Lemma 2.1. *Let X be an integer random variable X with third moment. Then*[2]

$$\mathrm{E}[|X - \mathrm{E}[X]|^3] \geq \mathrm{Var}[X]/2.$$

Proof. Let $\mu = \mathrm{E}[X]$. We may clearly assume that $0 \leq \mu < 1$ by replacing X with $X - a$ for an appropriate integer a. Define $Z_0 = [X \leq 0]$ and $Z_1 = [X \geq 1] = 1 - Z_0$. Then,

$$0 = \mathrm{E}[X - \mu] = \mathrm{E}[Z_1 \cdot |X - \mu|] - \mathrm{E}[Z_0 \cdot |X - \mu|],$$

and,

$$\mathrm{E}[|X - \mu|] = \mathrm{E}[Z_1 \cdot |X - \mu|] + \mathrm{E}[Z_0 \cdot |X - \mu|].$$

It follows that

$$\mathrm{Var}[X] = \mathrm{E}[Z_1 \cdot (X - \mu)^2] + \mathrm{E}[Z_0 \cdot (X - \mu)^2]$$

$$\geq (1 - \mu)\,\mathrm{E}[Z_1 \cdot |X - \mu|] + \mu\,\mathrm{E}[Z_0 \cdot |X - \mu|] = \frac{\mathrm{E}[|X - \mu|]}{2}. \tag{2.1}$$

Now finally,

$$\mathrm{Var}[X]^2 \leq \mathrm{E}[|X - \mu|] \cdot \mathrm{E}[|X - \mu|^3] \leq 2\,\mathrm{Var}[X]\,\mathrm{E}[|X - \mu|^3],$$

where the first inequality is by Cauchy-Schwartz and the second is an application of (2.1). The desired result follows. □

---

[2]Originally, our bound was $\mathrm{E}[|X - \mathrm{E}[X]|^3] \geq \mathrm{Var}[X]/10$, which sufficed for our purposes. We thank Ahmad Beirami [2] for pointing out how we could replace the constant 10 with 2 which is optimal as can be seen by letting X be a Bernoulli variable with parameter $1/2$.

Proof of Theorem 1.1. We start by noting that we may assume that $\sigma^2 > C$ for a sufficiently large constant C. Indeed, by Lemma 2.1,

$$c\left(\frac{\sum_{j\in[k]}\mathrm{E}[|X_j-\mu_j|^3]}{\sigma^3}\right)^2 \geq \frac{c}{4\sigma^2},$$

so (1.1) is immediate when $\sigma^2 \leq C$ (by choosing c sufficiently large).

Now the proof proceeds, similarly to proofs of the Berry-Esseen theorem and uses simple properties of the Fourier transformation of X. Let f_j be the characteristic function of $X_j - \mu_j$ for $j \in [k]$ and let F be the characteristic function of $X - \mu$. Then

$$F(\lambda) = \prod_{j\in[k]} f_j(\lambda) = \sum_{n\in\mathbb{Z}} \Pr[X=n]\,e^{i(n-\mu)\lambda}.$$

For non-zero integers s, it holds that $\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{is\lambda}d\lambda = 0$ whereas $\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{is\lambda}d\lambda = 1$ if $s = 0$. It follows that for any integer $n \in \mathbb{Z}$,

$$\Pr[X=n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(\lambda)e^{-i(n-\mu)\lambda}d\lambda.$$

In particular, if $\mu + t\sigma$ is an integer, then

$$\Pr[X=\mu+t\sigma] = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(\lambda)e^{-it\sigma\lambda}d\lambda.$$

We define $\tau = \frac{\sum_{j\in[k]}\mathrm{E}[|X_j-\mu_j|^3]}{\sigma^3}$ noting that we may assume that $\tau \leq c_0$ for a sufficiently small constant c_0 as otherwise the result is trivial. Split the interval $[-\pi,\pi]$ into three parts, $I_1 = [-\varepsilon,\varepsilon]$, $I_2 = [\varepsilon,\pi]$, and $I_3 = [-\pi,-\varepsilon]$. We will prove that if $\varepsilon = \frac{\sqrt{8\log 1/\tau}}{\sigma}$, then

$$\left|F(\lambda)e^{-it\sigma\lambda}\right| = |F(\lambda)| \leq \tau^2 \quad \text{for all } \lambda \in I_2 \cup I_3 \text{ and,} \tag{2.2}$$

$$\int_{-\varepsilon}^{\varepsilon} F(\lambda)e^{-it\sigma\lambda}d\lambda = \frac{1}{\sqrt{2\pi}\sigma}e^{-t^2/2} + O(\tau^2). \tag{2.3}$$

We note that we may assume that $\varepsilon \leq \pi$. Indeed, by Lemma 2.1, $\frac{\log 1/\tau}{\sigma^2} \leq \frac{\log(8\sigma)}{\sigma}$, and σ is assumed to be sufficiently large. The desired result thus follows immediately from (2.2) and (2.3).

We start by proving (2.2). It is a general fact that the characteristic function f_Y of a random variable Y is Hermitian, i.e., $f_Y(-t) = \overline{f_Y(t)}$. In particular, $|F(\lambda)| = |F(-\lambda)|$, so it is enough to prove that $|F(\lambda)| \leq \tau^2$ for $\lambda \in I_2$. As $|F|$ is decreasing on $[0,\pi]$, it in fact suffices to prove that $|F(\varepsilon)| \leq \tau^2$. Now another standard fact about the characteristic function f_Y of a random variable Y is that for any n,

$$\left|f_Y(\lambda) - \sum_{j=0}^{n}\frac{(i\lambda)^j}{j!}\mathrm{E}[Y^j]\right| \leq \frac{|\lambda|^{n+1}\mathrm{E}[|Y|^{n+1}]}{(n+1)!}. \tag{2.4}$$

By Jensen's inequality, $\sigma_j^2 \leq (\mathrm{E}[|X_j-\mu_j|^3])^{2/3}$, so it follows that

$$\varepsilon^2\sigma_j^2 \leq \left(\varepsilon^3\sum_{j\in[k]}\mathrm{E}\left[|X_j-\mu_j|^3\right]\right)^{2/3} = (\varepsilon^3\sigma^3\tau)^{2/3} \leq 8\log(1/\tau)\tau^{2/3} \leq 1, \tag{2.5}$$

where the last inequality used that $\tau \le c_0$ for a sufficiently small constant c_0. We may thus apply (2.4) with $n = 2$ to conclude that

$$|f_j(\varepsilon)| \le 1 - \frac{\varepsilon^2}{2}\sigma_j^2 + \mathrm{E}\left[|X_j - \mu_j|^3\right]\frac{\varepsilon^3}{6} \le e^{-\frac{\varepsilon^2}{2}\sigma_j^2 + \mathrm{E}\left[|X_j - \mu_j|^3\right]\frac{\varepsilon^3}{6}}.$$

Thus, for $\lambda \in I_2$,

$$\left|F(\lambda)e^{-it\sigma\lambda}\right| = |F(\lambda)| \le |F(\varepsilon)| \le e^{-\frac{\varepsilon^2}{2}\sigma^2 + \left(\sum_{j \in [k]} \mathrm{E}\left[|X_j - \mu_j|^3\right]\right)\frac{\varepsilon^3}{6}} = e^{-\varepsilon^2\sigma^2(1/2 - \sigma\varepsilon\tau/6)}.$$

As $\sigma\varepsilon\tau = \tau\sqrt{8\log 1/\tau} \le 3/2$, it therefore follows that for $\lambda \in I_2$,

$$|F(\lambda)| \le e^{-\frac{\varepsilon^2\sigma^2}{4}} = \tau^2, \tag{2.6}$$

which proves (2.2).

Turning to (2.3), we again use the Taylor expansion formula to get

$$f_j(\lambda) = 1 - \frac{\lambda^2}{2}\sigma_j^2 + \mathrm{E}\left[|X_j - \mu_j|^3\right]\lambda^3 g_j(\lambda),$$

for some (complex-valued) function $g_j(\lambda)$ with $|g_j(\lambda)| \le 1/6$ for all λ. As in (2.5), for $|\lambda| \le \varepsilon$,

$$\mathrm{E}\left[|X_j - \mu_j|^3\right]|\lambda|^3 \le 1, \tag{2.7}$$

and

$$\lambda^4\sigma_j^4 \le (|\lambda|^3\mathrm{E}[|X_j - \mu_j|^3])^{4/3} \le |\lambda|^3\mathrm{E}[|X_j - \mu_j|^3] \le 1. \tag{2.8}$$

It follows that $|f_j(\lambda) - 1| \le 5/6$. Now for $z \in \mathbb{C}$ with $|z| \le 5/6$ it holds that $1 + z = \exp(z + O(z^2))$. Also, if $a, b \in \mathbb{C}$ satisfy that $|a|^2 \le |b| \le 1$, then $|a+b|^2 \le 2(|a|^2 + |b|^2) \le 4|b|$. Combining these observations with (2.7) and (2.8) we find that

$$f_j(\lambda) = e^{-\frac{\lambda^2}{2}\sigma_j^2 + \mathrm{E}\left[|X_j - \mu_j|^3\right]O(\lambda^3)}.$$

It follows that

$$F(\lambda)e^{-it\sigma\lambda} = e^{-\frac{\lambda^2}{2}\sigma^2 + \left(\sum_{j \in [k]} \mathrm{E}\left[|X_j - \mu_j|^3\right]\right)O(\lambda^3)}e^{-it\sigma\lambda}$$

We then get that

$$F(\lambda)e^{-it\sigma\lambda} = e^{-\frac{\lambda^2}{2}\sigma^2 + \left(\sum_{j \in [k]} \mathrm{E}\left[|X_j - \mu_j|^3\right]\right)O(\lambda^3)}e^{-it\sigma\lambda}$$

$$= e^{-\frac{\lambda^2}{2}\sigma^2}\left(1 + \left(\sum_{j \in [k]} \mathrm{E}\left[|X_j - \mu_j|^3\right]\right)O(\lambda^3)\right)e^{-it\sigma\lambda} \tag{2.9}$$

for $|\lambda| \le \varepsilon$. Now we get that

$$\frac{1}{2\pi}\int_{-\varepsilon}^{\varepsilon}\left|e^{-\frac{\lambda^2}{2}\sigma^2}\left(\sum_{j \in [k]} \mathrm{E}\left[|X_j - \mu_j|^3\right]\right)O(\lambda^3)e^{-it\sigma\lambda}\right|d\lambda$$

$$= \frac{1}{\pi}\left(\sum_{j \in [k]} \mathrm{E}\left[|X_j - \mu_j|^3\right]\right)\int_0^{\varepsilon}e^{-\frac{\lambda^2}{2}\sigma^2}O(\lambda^3)d\lambda$$

$$= \frac{1}{\pi}\left(\frac{\sum_{j \in [k]} \mathrm{E}\left[|X_j - \mu_j|^3\right]}{\sigma^4}\right)\int_0^{\sqrt{8\log 1/\tau}}e^{-\frac{s^2}{2}}O(s^3)ds \tag{2.10}$$

$$= O\left(\frac{\sum_{j \in [k]} \mathrm{E}\left[|X_j - \mu_j|^3\right]}{\sigma^4}\right)$$

$$= O(\tau^2)$$

Here we used the substitution $s = \lambda\sigma$, and the last step uses Lemma 2.1. Again using the same substitution we get that

$$\frac{1}{2\pi}\int_{-\varepsilon}^{\varepsilon} e^{-\frac{\lambda^2}{2}\sigma^2} e^{-it\sigma\lambda} d\lambda = \frac{1}{2\pi\sigma}\int_{-\sqrt{8\log 1/\tau}}^{\sqrt{8\log 1/\tau}} e^{-\frac{s^2}{2}} e^{-its} ds$$

Note that for any $u > 0$,

$$\int_u^\infty e^{-s^2/2}\, ds \leq \frac{1}{u}\int_u^\infty s \cdot e^{-s^2/2}\, ds = \frac{1}{u}e^{-u^2/2},$$

so we can bound

$$\frac{1}{2\pi\sigma}\int_{|s|\geq\sqrt{8\log 1/\tau}} \left| e^{-\frac{s^2}{2}} e^{-its} \right| ds \leq \frac{\tau^4}{\pi\sigma\sqrt{8\log 1/\tau}} = O(\tau^2). \tag{2.11}$$

Calculating the Fourier transform of function of $e^{-\frac{s^2}{2}}$ we get that

$$\frac{1}{2\pi\sigma}\int_{-\infty}^{\infty} e^{-\frac{s^2}{2}} e^{-its} ds = \frac{1}{\sqrt{2\pi}\sigma}e^{-t^2/2} \tag{2.12}$$

Combining (2.9), (2.10), (2.11), and (2.12) proves (2.3). This finishes the proof. $\qquad\square$

3 Point Probabilities in the Tail

As is, Theorem 1.1 is only useful when $|t\sigma|$ is not too large. Indeed, for large $|t|$, the term $\frac{1}{\sqrt{2\pi}\sigma}e^{-t^2/2}$ will typically be much smaller than the error term on the right hand side of (1.1). We now show that if our variables satisfy the stronger property of being strongly monotone, we may also obtain precise estimates for the point probabilities in the tail by combining with the trick of exponential tilting.

Recall that we defined a real random variable X to be strongly monotone if $I_X \neq \{0\}$ and X_θ is monotone for each $\theta \in I_X$. Here, $I_X = \{\theta \in \mathbb{R} : \mathrm{E}[e^{\theta X}] < \infty\}$ consisted of those θ for which the moment generating function of X is defined, and X_θ was the exponentially tilted random variable defined by $\Pr[X_\theta = t] = \frac{\Pr[X=t]e^{\theta t}}{\mathrm{E}[e^{\theta X}]}$ for $t \in \mathbb{Z}$.

Many commonly encountered random variables have the property of being strongly monotone:

Lemma 3.1. *Let X be Bernoulli, Y be geometric and Z be Poisson. Then X, Y and Z are each strongly monotone.*

Proof. We already saw that the classes of Bernoulli, geometric, and Poisson variables consists of monotone variables. The result follows by calculating the point probabilities of the tilted variables (when they exists) and observing that each class is closed under exponential tilts. $\qquad\square$

Now suppose $X = \sum_{j\in[k]} X_j$ is a sum of independent random integer variables and moreover that X is not almost surely equal to a constant. We are interested in estimates for the probability $\Pr[X = t]$ for some $t \in \mathbb{Z}$. Let $I_j = \{\theta \in \mathbb{R} : \mathrm{E}[e^{\theta X_j}] < \infty\}$ and $I = \{\theta \in \mathbb{R} : \mathrm{E}[e^{\theta X}] < \infty\} = \cap_{j\in[k]} I_j$. We note each I_j and I are intervals containing 0. We define[3] $A = \mathrm{ess\,inf}\, X$ and $B = \mathrm{ess\,sup}\, X$. Let further $\psi_X : I \to \mathbb{R}$ be the cumulant generating function defined by $\psi_X : \theta \mapsto \log(\mathrm{E}[e^{\theta X}])$. It is well known that ψ_X is strictly convex and infinitely often differentiable for θ lying in the interior of I with

---

[3]Recall that the essential infimum and supremum of a random variable X are defined by $\mathrm{ess\,inf}\, X = \sup\{t : \Pr[X < t] = 0\}$ and $\mathrm{ess\,sup}\, X = \inf\{t : \Pr[X > t] = 0\}$ which are values in $\mathbb{R} \cup \{-\infty, \infty\}$.

$\psi'_X(\theta) = \frac{\mathrm{E}[Xe^{\theta X}]}{\mathrm{E}[e^{\theta X}]}$. For $t \in \mathbb{R}$, we define $g(t) = \sup_{\theta \in I}(\theta t - \psi_X(\theta))$. Now it is a standard fact about the cumulant generating function that if I contains a non-empty open interval (i.e., consists of more than a single point), then $\inf_{\theta \in I} \psi'_X(\theta) = A$ and $\sup_{\theta \in I} \psi'_X(\theta) = B$. If in particular $A < t < B$, there exists a θ_0 in the interior of I with $\psi'_X(\theta_0) = t$. Moreover, this θ_0 is unique since ψ_X is strictly convex.

Now let $(Y_j)_{j \in [k]}$ be independent random variables obtained by tilting each X_j by θ_0 as above. Let further $Y = \sum_{j \in [k]} Y_j$. For $s \in \mathbb{Z}$, we define $A_s = \{z \in \mathbb{Z}^k : z_1 + \cdots + z_k = s\}$. Then for any $t \in \mathbb{Z}$,

$$\Pr[X = t] = \sum_{z \in A_t} \prod_{j \in [k]} \Pr[X_j = z_j] = \frac{\mathrm{E}[e^{\theta_0 X}]}{e^{\theta_0 t}} \sum_{z \in A_t} \prod_{j \in [k]} \Pr[Y_j = z_j] = \frac{\mathrm{E}[e^{\theta_0 X}]}{e^{\theta_0 t}} \Pr[Y = t],$$

so Y is simply the variable obtained by tilting X by θ_0. Moreover, by the choice of θ_0,

$$\mathrm{E}[Y] = \sum_{z \in \mathbb{Z}} \frac{\Pr[X = z]e^{\theta_0 z} z}{\mathrm{E}[e^{\theta_0 X}]} = \frac{\mathrm{E}[Xe^{\theta_0 X}]}{\mathrm{E}[e^{\theta_0 X}]} = \psi'_X(\theta_0) = t.$$

Now the fact that $\mathrm{E}[Y] = t$, suggests using Theorem 1.1 to estimate the probability that $\Pr[Y = t]$. Doing so, we immediately obtain the following result.

Theorem 3.2. *Assume that X is strongly monotone and not almost surely equal to a constant. Moreover assume that $I \neq \{0\}$. Let t be an integer with with $A < t < B$ and θ be the unique real in the interior of I having $\psi'_X(\theta) = t$. Let Y be the exponential tilt of X by θ. Then $\mathrm{E}[Y] = t$ and*

$$\Pr[X = t] = \frac{\mathrm{E}[e^{\theta X}]}{e^{\theta t}} \left(\frac{1}{\sqrt{2\pi}\sigma_Y} \pm O\left(\frac{\eta_Y^2}{\sigma_Y^6} \right) \right), \tag{3.1}$$

where $\sigma_Y^2 = \mathrm{Var}[Y]$ and $\eta_Y = \sum_{j \in [k]} \mathrm{E}[|Y_j - \mathrm{E}[Y_j]|^3]$.

Remark 3.3. We note that if either $A = \mathrm{ess\,inf}\, X \neq -\infty$ or $B = \mathrm{ess\,sup}\, X \neq \infty$, then $[0, \infty) \subset I$ or $(-\infty, 0] \subset I$, respectively, and we can therefore always apply the exponential tilt in the lemma. We moreover note that for $t < A$ and $t > B$, it trivially holds that $\Pr[X = t] = 0$ and it is an easy exercise to show that

$$\Pr[X = A] = \prod_{j \in [k]} \Pr[X_j = \mathrm{ess\,inf}\, X_j], \quad \text{and} \quad \Pr[X = B] = \prod_{j \in [k]} \Pr[X_j = \mathrm{ess\,sup}\, X_j],$$

whenever $A \neq -\infty$ and $B \neq \infty$. Even though the lemma does not provide estimates for these probabilities, they are therefore usually easy to determine for concrete families of random variables.

To apply Theorem 3.2, for $X = \sum_{j \in [k]} X_j$ a concrete sum of strongly monotone random variables, say geometric variables, we would calculate ψ_X and find the unique θ with $\psi'_X(\theta) = t$. We would then determine the tilted random variables $(Y_j)_{j \in [k]}$. Typically Y_j comes from the same family of random variables as X_j, e.g., an exponential tilt of respectively a Bernoulli, geometric, and Poisson variable is again Bernoulli, geometric and Poisson. We would then determine the quantities η_Y and σ_Y^2 and plug into (3.1).

Example 3.4. Let $X = \sum_{j \in [k]} X_j$, where $(X_j)_{j \in [k]}$ are independent Bernoulli variables with $\Pr[X_j = 1] = p_j$. We want to estimate $\Pr[X = t]$ for some $0 < t < k - 1$. For this, we define θ, $(Y_j)_{j \in [k]}$ and Y as in Theorem 3.2. Then each Y_j is again Bernoulli. If $\Pr[Y_j = 1] = q_j$, then $\mathrm{E}[|Y_j - \mathrm{E}[Y_j]|^3] = q_j(1 - q_j)(q_j^2 + (1 - q_j)^2) \leq \mathrm{Var}[Y_j]$, so that $\eta_Y \leq \sigma_Y^2$. Thus, the bound of (3.1) becomes

$$\Pr[X = t] = \frac{1}{\sqrt{2\pi}\sigma_Y} \frac{\mathrm{E}[e^{\theta X}]}{e^{\theta t}} \left(1 \pm O\left(\frac{1}{\sigma_Y} \right) \right),$$

The bound on the error term can be shown to be asymptotically tight using known results. We in particular note that if $\sigma_Y = \omega(1)$, the bound on $\Pr[X = t]$ is within a factor of $1 + o(1)$ of the true value. Consider as a very simple example[4], the case where the Bernoulli variables $(X_j)_{j \in [k]}$ are identically distributed. Then the same holds for the $(Y_j)_{j \in [k]}$, and since $\mathrm{E}[Y] = t$, we must have that $\sigma_Y^2 = t(1 - t/k)$. In particular, the bound is within a factor of $1 + o(1)$ of the true value as long as $\omega(1) < t < k - \omega(1)$.

Example 3.5. Let $X = \sum_{j \in [k]} X_j$ be a sum of independent geometric variables such that for some probabilities $(p_j)_{j \in [k]}$ and each $s \in \mathbb{N}_0$, $\Pr[X_j = s] = p_j^s(1 - p_j)$. Let $\mu_j = \mathrm{E}[X_j] = \frac{p_j}{1 - p_j}$ for $j \in [k]$ and $\mu = \mathrm{E}[X]$. Assume that $\mu_j = O(1)$ for $j \in [k]$. We want to estimate $\Pr[X = t]$ for some integer $t > 0$ using Theorem 3.2, and we define θ, $(Y_j)_{j \in [k]}$ and Y accordingly. For simplicity, we will assume that $t = O(\mathrm{E}[X])$. By Lemma 3.1, each Y_j is again geometric, say with $\Pr[Y_j = s] = q_j^s(1 - q_j)$ for $s \in \mathbb{N}_0$. Moreover, since $\mathrm{E}[X_j] = O(1)$ for $j \in [k]$ and $t = O(\mathrm{E}[X])$, it follows that also $\mathrm{E}[Y_j] = O(1)$ for $j \in [k]$. Now simple calculations yields that $\mathrm{Var}[Y_j] = \Theta(\mathrm{E}[Y_j])$ and $\mathrm{E}[|Y_j - \mathrm{E}[Y_j]|^3] = \Theta(\mathrm{E}[Y_j])$. Plugging into (3.1), we thus obtain that

$$\Pr[X = t] = \frac{1}{\sqrt{2\pi}\sigma_Y} \frac{\mathrm{E}[e^{\theta X}]}{e^{\theta t}} \left(1 \pm O\left(\frac{1}{\sqrt{\mu_Y}}\right)\right),$$

where $\mu_Y = \mathrm{E}[Y] = t$. In particular, the bound is within a factor of $1 + o(1)$ of the true value as long as $t = \omega(1)$.

References

[1] Aamand, A., Knudsen, J. B. T., and Thorup, M. Load balancing with dynamic set of balls and bins. In *Proc. 53rd ACM SIGACT Symposium on Theory of Computing (STOC)* (2021), ACM, pp. 1262–1275. Full version on https://arxiv.org/abs/2104.05093.

[2] Beirami, A. $\mathbb{E}[|x - \mu|^n] \le 2\mathbb{E}[|x - \mu|^{n+1}]$ for integer random variables. Mathematics Stack Exchange. URL: https://math.stackexchange.com/q/4102849 (version: 2021-04-16).

[3] Blackwell, D., and Jr., J. L. H. The Probability in the Extreme Tail of a Convolution. *The Annals of Mathematical Statistics 30*, 4 (1959), 1113 – 1120.

[4] Dolgopyat, D., and Hafouta, Y. Edgeworth expansions for independent bounded integer valued random variables. *Stochastic Processes and their Applications 152* (2022), 486–532.

[5] Gnedenko, B. V., and Kolmogorov, A. N. Limit distributions for sums of independent random variables. *Addison-Wesley, Cambridge 121*, 3149 (1955).

[6] Ibragimov and Linnik, I. V. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Groningen, 1971.

[7] Iltis, M. Sharp asymptotics of large deviations in $\mathbb{R}^d$. *Journal of Theoretical Probability 8*, 3 (1995), 501–522.

[8] McDonald, D. A local limit theorem for large deviations of sums of independent, nonidentically distributed random variables. *Annals of Probability 7* (1979), 526–531.

---

[4]This example is to be seen as a proof of concept as one can obtain precise asymptotically tight estimates directly using Stirling's formula.