Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 121 (2011) 1138–1147

stochastic processes and their applications

www.elsevier.com/locate/spa

Conditional distribution of heavy tailed random variables on large deviations of their sum

Inés Armendáriz[a], Michail Loulakis[b,c,*]

[a] *Universidad de San Andrés, Vito Dumas 284, B1644BID, Victoria, Argentina*
[b] *Department of Applied Mathematics, University of Crete, Knossos Avenue, 714 09 Heraklion Crete, Greece*
[c] *Institute of Applied and Computational Mathematics, FORTH, Heraklion Crete, Greece*

Received 7 November 2010; received in revised form 21 January 2011; accepted 26 January 2011
Available online 12 February 2011

Dedicated to S.R.S. Varadhan on the occasion of his birthday

Abstract

It is known that large deviations of sums of subexponential random variables are most likely realised by deviations of a single random variable. In this article we give a detailed picture of how subexponential random variables are distributed when a large deviation of the sum is observed.

MSC: 60F10

Keywords: Large deviations; Subexponential distributions; Conditional limit theorem; Gibbs conditioning principle

1. Introduction

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common distribution μ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let

$$S_n = X_1 + \cdots + X_n, \quad n \geq 1.$$

The most classical problem in large deviations is establishing asymptotic expressions for

$$\bar{F}_n(x) := \mathbb{P}[S_n > x] \tag{1.1}$$

* Corresponding author at: Department of Applied Mathematics, University of Crete, Knossos Avenue, 714 09 Heraklion Crete, Greece. Tel.: +30 6972424164.

E-mail addresses: iarmendariz@udesa.edu.ar (I. Armendáriz), loulakis@tem.uoc.gr (M. Loulakis).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2011.01.011

when this quantity converges to zero. The answer depends heavily on the nature of the tails of the distribution μ. When the moment generating function is finite in a neighborhood of the origin (Cramér's condition) Cramér derived asymptotic expressions for $\bar{F}_n(x)$ valid uniformly over different ranges of x-values. These results were later refined by Petrov (cf. [15].) In this case Gibbs conditioning principle provides an answer to how a large deviation of the sum is typically realised: subject to the large deviation, the random variables $\{X_i\}$ become independent in the limit, but their marginal distribution is modified in such a way that the behavior imposed on the sum now becomes typical. In particular, no single random variable becomes excessively large compared to the others.

The situation is totally different when Cramér's condition is violated. It is known since the classical works of Heyde [10] and Nagaev [12] that large deviations of sums of independent heavy tailed random variables are typically realised by one random variable taking a very large value. In this article we investigate the conditional distribution of the random variables $\{X_i\}_{1\le i\le n}$ subject to a large deviation of their sum S_n. It turns out that as $n \to \infty$ this conditional distribution converges to a product of $n-1$ copies of μ, while the remaining variable realises the large deviation event by taking a very large value. We determine when the fluctuations around that value have a scaling limit, and we show that given the sum exceeds a large value, the maximum is asymptotically independent of the smallest variables.

2. Notation and results

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with common distribution μ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote by F their distribution function $F(x) = \mu(-\infty, x]$. We are interested in the case where F is in the class of subexponential distributions, that is

$$\lim_{x\to\infty} \frac{\bar{F}(x+y)}{\bar{F}(x)} = 1 \quad \forall y \in \mathbb{R}, \tag{2.1}$$

with $\bar{F}(x) = \mathbb{P}[X_k > x]$, and

$$\lim_{x\to\infty} \frac{\bar{F}_n(x)}{n\bar{F}(x)} = 1 \quad \forall n \in \mathbb{N}, \tag{2.2}$$

where $\bar{F}_n(x)$ is defined in (1.1). If the support of μ is contained in the positive half-line, then (2.1) is implied by (2.2), and in that case subexponentiality can be defined by the latter condition alone. Since it is generally true that for all $n \in \mathbb{N}$ we have

$$\lim_{x\to\infty} \frac{\mathbb{P}[\max_{1\le k\le n} X_k > x]}{n\bar{F}(x)} = 1.$$

Eq. (2.2) states that the tail of the sum in a sample of independent μ-distributed random variables is determined by the tail of the largest variable. These distributions arise naturally when modeling heavy tailed phenomena. For instance, individual claims in insurance or large interarrival times in queuing systems are usually modeled by distributions of this kind. Typical members of this class include distributions with regularly varying, lognormal-type, or Weibull-type tails. Sufficient conditions for a given distribution to be subexponential that are straightforward to check can be found in [14].

An immediate consequence of (2.2) is the existence of a sequence $d_n \to \infty$ such that

$$\lim_{n\to\infty} \sup_{x\geq d_n} \left| \frac{\bar{F}_n(x)}{n\bar{F}(x)} - 1 \right| = 0. \tag{2.3}$$

Since it is important in applications to have estimates for the threshold d_n for which (2.3) holds, a large amount of work has been done in this direction. Interested readers can find reviews on the topic in [13,15]. A very nice account is also provided by Mikosch and Nagaev in [11]. Denisov et al. give an up-to-date treatment of this problem in [5].

When the distribution μ satisfies a local version of (2.2), a local version of (2.3) is valid. Let $\Delta = (0, s]$ for some $s > 0$ and denote by $x + \Delta$ the interval $(x, x + s]$. We say that μ is Δ-subexponential if $\mu[x + \Delta] > 0$ for all sufficiently large x and

$$\lim_{x\to\infty} \frac{\mu[x + y + \Delta]}{\mu[x + \Delta]} = 1 \quad \forall y \in \mathbb{R}, \tag{2.4}$$

and

$$\lim_{x\to\infty} \frac{\mathbb{P}[S_n \in x + \Delta]}{n\mu[x + \Delta]} = 1 \quad \forall n \in \mathbb{N}. \tag{2.5}$$

The concept of Δ-subexponentiality was introduced in [2] by Asmussen et al. A Δ-subexponential distribution is also $m\Delta$-subexponential for all $m \in \mathbb{N}$, $m\Delta = (0, ms]$, and subexponential in the sense of (2.1) and (2.2) (cf. [2].) Even though there are examples of subexponential distributions that are not Δ-subexponential for finite Δ, most distributions that are used in practice are. This can be easily verified using the sufficient conditions for Δ-subexponentiality provided in [2]. The asymptotics for the large deviation probabilities are now given by

$$\lim_{n\to\infty} \sup_{x\geq d_n} \left| \frac{\mathbb{P}[S_n \in x + \Delta]}{n\mu[x + \Delta]} - 1 \right| = 0, \tag{2.6}$$

and sufficient conditions on d_n for (2.6) to hold can be found in [5].

We are interested in the conditional distribution of the variables $\{X_n\}$ subject to a large deviation of their sum. Assuming that (2.6) holds for some interval Δ that may be finite or infinite, we would like to determine the asymptotic behavior of

$$\mu^{\Delta}_{n,x}[A] = \mathbb{P}[(X_1, \ldots, X_n) \in A \mid S_n \in x + \Delta]$$

when $n \to \infty$ and $x \geq d_n$. Note that when $\Delta = (0, \infty)$ the definition of Δ-subexponentiality reduces to the standard definition of subexponentiality and (2.6) reduces to (2.3). This allows to treat rare events of the form $\{S_n \in (x, x + s]\}$ or $\{S_n > x\}$ simultaneously in Theorem 1 below.

A related question was raised in [9] where certain subexponential families μ of lattice type are considered under $\{S_n = x(n)\}$, and it is shown that the finite dimensional marginals of the conditional distribution converge to a product of copies of μ. In [8] the authors consider a family of discrete distributions μ that includes those with regularly varying tails, subject to $\{S_n = x\}$ where n is fixed and $x \to \infty$, and they show a version of (2.9) below. In this article we show that for all Δ-subexponential distributions μ, the conditional distribution of the $n - 1$ smallest variables subject to $\{S_n \in x + \Delta\}$ approaches in total variation a product of $n - 1$ copies of μ as $n \to \infty$, as long as $x \geq d_n \vee \ell_n$. The sequence ℓ_n is defined in (3.3), it can be easily computed from F, and in most interesting cases turns out to be smaller than d_n.

We will denote by $T : \cup_{n\in\mathbb{N}} \mathbb{R}^n \to \cup_{n\in\mathbb{N}} \mathbb{R}^n$ the operator that exchanges the last and the maximum component of a finite sequence:

$$T(x_1, \ldots, x_n)_k = \begin{cases} \max\limits_{1\le i\le n} x_i & \text{if } k = n, \\ x_n & \text{if } x_k > \max\limits_{1\le i<k} x_i \text{ and } x_k = \max\limits_{i\ge k} x_i, \\ x_k & \text{otherwise.} \end{cases}$$

Theorem 1. *Suppose μ is Δ-subexponential, take d_n as in (2.6) and ℓ_n as in (3.3). If $q_n = d_n \vee \ell_n$, then*

$$\lim_{y\to\infty} \sup_{n\in\mathbb{N}} \sup_{x\ge q_n\vee y} \sup_{A\in\mathcal{B}(\mathbb{R}^{n-1})} |\mu^{\Delta}_{n,x} \circ T^{-1}[A \times \mathbb{R}] - \mu^{n-1}[A]| = 0. \tag{2.7}$$

Remarks. 1. Since $q_n \to \infty$, (2.7) is equivalent to the following two conditions

$$\lim_{n\to\infty} \sup_{x\ge q_n} \sup_{A\in\mathcal{B}(\mathbb{R}^{n-1})} |\mu^{\Delta}_{n,x} \circ T^{-1}[A \times \mathbb{R}] - \mu^{n-1}[A]| = 0, \tag{2.8}$$

and

$$\lim_{x\to\infty} \sup_{A\in\mathcal{B}(\mathbb{R}^{n-1})} |\mu^{\Delta}_{n,x} \circ T^{-1}[A \times \mathbb{R}] - \mu^{n-1}[A]| = 0, \quad \forall n \in \mathbb{N}. \tag{2.9}$$

Even though (2.9) alone implies (2.8) for some sequence q_n, Theorem 1 also contains information about the threshold q_n in (2.8) that can be useful for applications.

2. Note that $\mu^{\Delta}_{n,x} \circ T^{-1}[A \times \mathbb{R}]$ is the measure assigned to $A \in \mathbb{R}^{n-1}$ by the conditional distribution of the $n-1$ smallest variables. In other words, Theorem 1 states that under (2.3), conditioning on $\{S_n \in x + \Delta\}$ affects only the maximum in the limit, and the $n-1$ smallest variables become asymptotically independent. Such a result is rather uncommon, and when μ satisfies Cramér's condition an analogous statement is not true. Now, any limit theorem for i.i.d. random variables with distribution μ can be cast in this setting. For instance, we could obtain conditional limit theorems for the statistics of any order $k > 1$: the k-th order statistic of $(X_1, \ldots, X_n)$ subject to the condition $S_n \in x + \Delta$, $x \ge q_n$, asymptotically behaves like the $(k-1)$-th order statistic of an independent sample.

Unlike the asymptotic independence of the smallest variables, the fluctuations of the maximum $M_n = \max_{1\le i\le n} X_i$ and its dependence on the smallest variables are influenced by the form of conditioning. When $\Delta = (0, s]$ the condition we impose on the sum is very restrictive and the fluctuations of the maximum are determined by the fluctuations of the sum of the smallest variables. This can be easily seen since $\mu^{\Delta}_{n,x}[M_n + \sum_{j=1}^{n-1}(TX)_i \in (x, x+s]\,] = 1$ by definition. Therefore, if the (unconditioned) distribution of S_{n-1}/b_n converges to a stable law H, it follows immediately from Theorem 1 that under $\mu^{\Delta}_{n,x}$ we have

$$\frac{M_n - x}{b_n} \xrightarrow{d} -H. \tag{2.10}$$

Note that the converse is also true. In particular, the fluctuations of the conditional maximum are typically two-sided and they have a nontrivial scaling limit if and only if μ is attracted to a stable distribution. In [1] Theorem 1 is proved for a particular family of lattice distributions subject to $\{S_n = x\}$, and this observation is used to obtain a limit theorem for the fluctuations of the maximum in a system of interacting particles.

On the other hand when we condition on $\{S_n > x\}$ it turns out that the maximum coordinate is asymptotically independent of the smallest variables, its fluctuations around x are one-sided, and they have a nontrivial scaling limit if and only if μ is in the maximum domain of attraction of an extreme value distribution. For ease of notation, we will now drop $\Delta = (0, \infty)$ from the notation,

$$\mu_{n,x}[A] = \mathbb{P}[(X_1, \dots, X_n) \in A \mid S_n > x].$$

Let ν_x stand for the conditional distribution of X_i subject to $X_i > x$. That is,

$$\nu_x[A] = \mathbb{P}[X_i \in A \mid X_i > x] = \frac{\mu[A \cap (x, \infty)]}{\bar{F}(x)}.$$

We will use $\|\nu\|_{t.v.}$ to denote the total variation norm of a signed Borel measure on $\mathbb{R}^n$. That is

$$\|\nu\|_{t.v.} = \sup_{A \in \mathcal{B}(\mathbb{R}^n)} |\nu(A)|.$$

Theorem 2. *Suppose μ is subexponential. Then*

$$\lim_{y\to\infty} \sup_{n\in\mathbb{N}} \sup_{x \geq q_n \vee y} \|\mu_{n,x} \circ T^{-1} - (\mu^{n-1} \times \nu_x)\|_{t.v.} = 0, \tag{2.11}$$

where q_n is the sequence appearing in Theorem 1.

Since the distribution of $(X_1, \dots, X_n)$ subject to $\{S_n > x\}$ is clearly exchangeable, the position of the maximum coordinate is uniformly distributed among $1, \dots, n$. Theorem 2 states that the conditional distribution of the maximum coordinate becomes asymptotically a randomly located ν_x, while the law of the remaining $n - 1$ variables is the product μ^{n-1} as was established in Theorem 1.

It is interesting to examine whether (2.11) entails a limit theorem for the fluctuations of the maximum around x, that is, whether there exists a scaling function $\psi(\cdot)$ such that under $\mu_{n,x}$ we have

$$\frac{M_n - x}{\psi(x)} \xrightarrow{d} \Lambda, \tag{2.12}$$

for some nontrivial distribution Λ. In view of Theorem 2 this is equivalent to asking when

$$\nu_x[(x + u\psi(x), \infty)] = \frac{\bar{F}(x + u\psi(x))}{\bar{F}(x)} \tag{2.13}$$

converges as $x \to \infty$ to a nontrivial function of u. This is precisely the subject of [3], where Balkema and de Haan determine all possible scaling limits of residual life times as the survival time goes to infinity, and the corresponding domains of attraction. It follows from their results (Theorems 1, 3 and 4 there) that nontrivial limits in the right hand side of (2.12) can only be of two types.

1. An exponential distribution of rate 1 if and only if μ is in the maximum domain of attraction of the Gumbel distribution. In this case ψ can be determined by requiring the expression in (2.13) to converge to e^{-u}.

2. A Pareto distribution on $\mathbb{R}_+$ with $\bar{\Lambda}(u) = (1+u)^{-\alpha}$ and $\alpha > 0$, if and only if μ has regularly varying tails with index $-\alpha$, that is $\bar{F}(x) = x^{-\alpha}L(x)$ as $x \to \infty$, and L is a slowly varying function. Note that this is equivalent to μ being in the maximum domain of attraction of the Fréchet distribution with index α (cf. [4].) In this case $\psi(x) = x$.

That regularly varying distributions satisfy our main assumption (2.3) is a long known fact (cf. [10,12].) In particular, if $\alpha > 2$ one can choose $d_n = \sqrt{tn \log n}$ for any $t > \alpha - 2$ (cf. [13].) The articles [5,11] are excellent references for subexponential distributions in the maximum domain of attraction of the Gumbel distribution and the corresponding sequences d_n for which (2.3) holds.

Remarks. 1. Theorem 2 and the discussion following it generalise a result of Mikosch and Nagaev (Proposition 4.4 in [11]) where they prove (2.12) for some of the most commonly used subexponential distributions.

2. Theorem 2 also generalises an old result by Richard Durrett [6]. In that article it is proved that if μ is regularly varying with index $\alpha < -2$ and $\mathbb{E}[X_1] = -b < 0$, then

$$\left(\frac{S_{[n\cdot]}}{n} \,\middle|\, S_n > 0 \right) \Rightarrow J_{\alpha,b} 1_{\{U \le \cdot\}} - b\,\cdot,$$

where U is uniform in [0,1] and $J_{\alpha,b}$ is independent of U with Pareto distribution. As in Corollary 3 in [1], Theorem 2 also establishes a conditional invariance principle for the sum of the random variables cut-off at a level εn.

The rest of the article is devoted to the proof of the theorems.

3. Proof of the theorems

Given a vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we denote by $M_{\mathbf{x}}$ the coordinate of maximum size and by $m_{\mathbf{x}}$ its position. Precisely,

$$M_{\mathbf{x}} = \max_{1 \le k \le n} x_i \quad \text{and} \quad m_{\mathbf{x}} = k \Leftrightarrow x_k > x_j, \quad j < k, \text{ and } x_k \ge x_j, \; j \ge k.$$

We will denote by σ^j the operator that exchanges the j-th and the last coordinate of $\mathbf{x}$, that is

$$\sigma^j(x_1, \ldots, x_j, \ldots, x_{n-1}, x_n) = (x_1, \ldots, x_n, \ldots, x_{n-1}, x_j).$$

With this notation we may write $T\mathbf{x} = \sigma^{m_{\mathbf{x}}}\mathbf{x}$. Let also $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{X}^{n-1} = (X_1, \ldots, X_{n-1})$. We begin with a proof to Theorem 2 that only works when μ is supported on the positive half-line, but has the advantage of being intuitive and straightforward.

Proof of Theorem 2 when μ is supported on $[0, \infty)$**.** Observe that if m is a probability measure on a σ-field $\mathcal{F}$, $A \in \mathcal{F}$ is such that $m[A] > 0$, and m_A is the measure m conditioned on A, that is $m_A[B] = m[B \mid A]$, then m_A is the solution to the minimisation problem

$$\min_{\nu[A]=1} H(\nu|m), \tag{3.1}$$

where the supremum is taken over all probability measures on $\mathcal{F}$ supported in A, and $H(\nu|m)$ is the relative entropy of ν with respect to m, i.e. $H(\nu|m) = \int f \log f dm$ if $\nu \ll m$ with $f = \mathrm{d}\nu/dm$, and $H(\nu|m) = \infty$ otherwise. To see this, integrate over A the elementary inequality $f \log f \ge cf - e^{c-1}$ and optimise for $c \in \mathbb{R}$.

By Csiszár's parallelogram identity

$$H(\pi|m) + H(\nu|m) = 2H\left(\frac{\pi+\nu}{2}\middle| m\right) + H\left(\pi \middle| \frac{\pi+\nu}{2}\right) + H\left(\nu \middle| \frac{\pi+\nu}{2}\right)$$

and Pinsker's inequality $H(\pi|\nu) \geq 2\|\pi - \nu\|^2_{t.v.}$, we get

$$H(\pi|m) + H(\nu|m) - 2H\left(\frac{\pi+\nu}{2}\middle| m\right) \geq \|\pi - \nu\|^2_{t.v.}.$$

Hence, if π solves (3.1) and $\nu[A] = 1$ we have $\|\pi - \nu\|^2_{t.v.} \leq H(\nu|m) - H(\pi|m)$.

Now take $A = \{\sum x_i > x\}$, $m = \mu^n$, $\pi = \mu_{n,x}$, and $\nu = \mu^*_{n,x} = \frac{1}{n}\sum_{j=1}^n \sigma^j(\mu^{n-1} \times \nu_x)$. Note that $\mu^*_{n,x}[S_n > x] = 1$, since μ is supported on $[0,\infty)$ and $\nu_x(x,\infty) = 1$. The density of $\mu^*_{n,x}$ with respect to μ^n is given by

$$f_n(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n} \frac{\mathbb{1}\{x_j > x\}}{\bar{F}(x)} = \frac{N_n(\mathbf{x})}{n\bar{F}(x)},$$

where $N_n(\mathbf{x})$ stands for the number of coordinates in $\mathbf{x}$ that are greater than x. Hence,

$$\|\mu_{n,x} - \mu^*_{n,x}\|^2_{t.v.} \leq H(\mu^*_{n,x}|\mu^n) - H(\mu_{n,x}|\mu^n) = \log\left(\frac{\bar{F}_n(x)}{n\bar{F}(x)}\right) + \int \log N_n(\mathbf{x})\, \mathrm{d}\mu^*_{n,x}.$$

The last integral can be estimated as follows.

$$\begin{aligned}\int \log N_n(\mathbf{x})\, \mathrm{d}\mu^*_{n,x} &= \int \log N_n(\mathbf{x})\, \mathrm{d}\mu^{n-1}\mathrm{d}\nu_x = \int \log\left(1 + N_{n-1}(\mathbf{x}^{n-1})\right)\, \mathrm{d}\mu^{n-1} \\ &\leq \int N_{n-1}(\mathbf{x}^{n-1})\mathrm{d}\mu^{n-1} = (n-1)\bar{F}(x).\end{aligned}$$

If we let $x \to \infty$, or if we keep $x \geq d_n$ (with $n\bar{F}(d_n) \to 0$) and let $n \to \infty$, we have $\|\mu_{n,x} - \mu^*_{n,x}\|_{t.v.} \to 0$ by (2.2) or (2.3), respectively. □

We continue now with some elementary observations that will be useful for both proofs.

The convergence in (2.4) is in fact uniform over compact y-sets. This follows from the uniform convergence theorem for slowly varying functions (see [4], Theorem 1.2.1), as (2.4) implies that $x \mapsto \mu[\log x + \Delta]$ is slowly varying. In particular, if b_n is any sequence growing to infinity there exists a sequence $m_n \to \infty$ such that

$$\lim_{n\to\infty} \sup_{x\geq m_n} \sup_{0\leq y\leq b_n} \left|\frac{\mu[x-y+\Delta]}{\mu[x+\Delta]} - 1\right| = 0. \tag{3.2}$$

Take a sequence b_n such that S_{n-1}/b_n is tight and choose a sequence $\ell_n \gg b_n$ such that

$$D_n(L) := \sup_{x\geq \ell_n} \sup_{|y|\leq Lb_n} \left(1 - \frac{\mu[x-y+\Delta]}{\mu[x+\Delta]}\right) \longrightarrow 0 \quad \text{as } n\to\infty,\ \forall L > 0. \tag{3.3}$$

To see that such a sequence exists, iterate (3.2) using the fact the limit is uniform in $x \geq m_n$ to get

$$\lim_{n\to\infty} \sup_{x\geq m_n} \sup_{-Lb_n\leq y\leq b_n} \left|\frac{\mu[x-y+\Delta]}{\mu[x+\Delta]} - 1\right| = 0. \tag{3.4}$$

Now, if ρ_n is any sequence increasing to infinity we may choose $\ell_n = m_n + \rho_n b_n$.

The sequence q_n in the statement of the theorems is chosen as $q_n = d_n \vee \ell_n$. Very often, in fact in all cases we are aware of where a threshold d_n in (2.3) or (2.6) is explicitly known, and certainly for the d_n constructed in [5], we can choose $\ell_n \le d_n$ so the supremum in the Theorems is taken for $x \ge d_n$.

Finally, note that for any $B \in \mathcal{B}(\mathbb{R}^n)$ we have

$$\begin{aligned}\mathbb{P}[T\mathbf{X} \in B, S_n \in x + \Delta] &= \sum_{j=1}^{n} \mathbb{P}[T\mathbf{X} \in B, S_n \in x + \Delta,\ m_{\mathbf{X}} = j] \\ &\ge \sum_{j=1}^{n} \mathbb{P}[\sigma^j \mathbf{X} \in B, S_n \in x + \Delta,\ m_{\sigma^j \mathbf{X}} = n] \\ &= n\,\mathbb{P}[\mathbf{X} \in B, S_n \in x + \Delta,\ m_{\mathbf{X}} = n]. \end{aligned} \tag{3.5}$$

The last equality holds because $\mathbb{P}$ is invariant under σ^j. The penultimate inequality holds because $m_{\sigma^j \mathbf{x}} = n \Rightarrow m_{\mathbf{x}} = j$ (notice however that if μ is atomless this inequality and (3.6) below are in fact equalities.) In view of (3.5) we have

$$\mu_{n,x}^{\Delta} \circ T^{-1}[B] \ge \frac{n\,\mathbb{P}[\mathbf{X} \in B,\ S_n \in x + \Delta,\ m_{\mathbf{X}} = n]}{\mathbb{P}[S_n \in x + \Delta]}. \tag{3.6}$$

Proof of Theorem 1. Consider $A \in \mathcal{B}(\mathbb{R}^{n-1})$ as in the statement of the theorem and fix $L \in \mathbb{N}$. We have

$$\begin{aligned}&\mathbb{P}[\mathbf{X} \in A \times \mathbb{R},\ S_n \in x + \Delta,\ m_{\mathbf{X}} = n] \\ &\quad \ge \mathbb{P}[\mathbf{X} \in A \times \mathbb{R},\ S_n \in x + \Delta,\ |S_{n-1}| < Lb_n,\ m_{\mathbf{X}} = n] \\ &\quad \ge \mathbb{P}[\mathbf{X} \in A \times \mathbb{R},\ S_n \in x + \Delta,\ |S_{n-1}| < Lb_n,\ M_{\mathbf{X}^{n-1}} \le x - Lb_n] \\ &\quad = \int_{\mathbf{X}^{n-1} \in A \cap G} \mu[x - S_{n-1} + \Delta]\,\mathrm{d}\mathbb{P},\end{aligned}$$

where

$$G = G(n, L, x) = \left\{ \mathbf{u} \in \mathbb{R}^{n-1} : \left| \sum_{i=1}^{n-1} u_i \right| < Lb_n,\ M_{\mathbf{u}} \le x - Lb_n \right\}. \tag{3.7}$$

Notice that when $\mathbf{u} \in G$ and $x \ge \ell_n$ we have

$$\mu\left[x - \sum_{i=1}^{n-1} u_i + \Delta \right] \ge (1 - D_n(L))\mu[x + \Delta],$$

so that (3.6) can be reinforced to

$$\mu_{n,x}^{\Delta} \circ T^{-1}[A \times \mathbb{R}] \ge (1 - D_n(L))\, \frac{n\mu[x + \Delta]}{\mathbb{P}[S_n \in x + \Delta]}\, \mathbb{P}[\mathbf{X}^{n-1} \in A \cap G]$$

giving the estimate

$$\begin{aligned}&\mu_{n,x}^{\Delta} \circ T^{-1}[A \times \mathbb{R}] - \mathbb{P}[\mathbf{X}^{n-1} \in A] \\ &\quad \ge -\left(\mathbb{P}[\mathbf{X}^{n-1} \notin G] + D_n(L) + \left| \frac{n\mu[x + \Delta]}{\mathbb{P}[S_n \in x + \Delta]} - 1 \right| \right).\end{aligned}$$

Denote the expression in the parenthesis on the right hand side above by $R(n, L, x)$. We can get an upper bound by applying the same estimate for $\mathbb{R}^{n-1} \setminus A$, the complement of A. Combining the two bounds we get

$$|\mu^{\Delta}_{n,x} \circ T^{-1}[A \times \mathbb{R}] - \mathbb{P}[\mathbf{X}^{n-1} \in A]| \le R(n, L, x).$$

Now the sequence S_{n-1}/b_n is tight, so we have

$$\lim_{L\to\infty} \sup_n \mathbb{P}[|S_{n-1}| \ge Lb_n] = 0. \tag{3.8}$$

On the other hand, it is known (see [7], Section lX.7) that

$$\lim_{L\to\infty} \sup_n n[F(-Lb_n) + \bar{F}(Lb_n)] = 0, \tag{3.9}$$

and since $\ell_n \gg b_n$ we have

$$\sup_{x\ge\ell_n} \mathbb{P}[M_{\mathbf{X}^{n-1}} > x - Lb_n] = 1 - \inf_{x\ge\ell_n} \left(1 - \bar{F}(x - Lb_n)\right)^{n-1} \longrightarrow 0 \quad \text{as } n \to \infty.$$

Combining this limit with (3.8) we see that $\mathbb{P}[\mathbf{X}^{n-1} \notin G]$ vanishes uniformly on $x \ge \ell_n$ as $n \to \infty$ and then $L \to \infty$. Relation (2.8) now follows from (2.6) and (3.3). Relation (2.9) is proved similarly, using (2.4) and (2.5). □

Proof of Theorem 2. The proof follows the general outline of Theorem 1.

It is sufficient to show that

$$\lim_{n\to\infty} \sup_{x\ge d_n\vee\ell_n} \sup_{U\in\mathcal{R}} |\mu_{n,x} \circ T^{-1}[U] - \mu^{n-1} \times \nu_x[U]| = 0, \tag{3.10}$$

where the supremum above is taken over the class $\mathcal{R}$ of finite disjoint unions of rectangles $A_j \times B_j$ with $A_j \in \mathcal{B}(\mathbb{R}^{n-1})$ and $B_j \in \mathcal{B}(\mathbb{R})$. Recall the definition of $G \subset \mathbb{R}^{n-1}$ in (3.7) and define $I = (x + Lb_n, \infty)$. Using (3.6) we have

$$\begin{aligned}
\mu_{n,x} \circ T^{-1}\left[\bigcup_j A_j \times B_j\right] &= \sum_j \mu_{n,x} \circ T^{-1}[A_j \times B_j] \\
&\ge \sum_j \frac{n\mathbb{P}[\mathbf{X}^{n-1} \in A_j,\ X_n \in B_j,\ S_n > x,\ m_{\mathbf{X}} = n]}{\bar{F}_n(x)} \\
&\ge \frac{n}{\bar{F}_n(x)} \sum_j \mathbb{P}[\mathbf{X}^{n-1} \in A_j \cap G,\ X_n \in B_j \cap I] \\
&= \frac{n\bar{F}(x)}{\bar{F}_n(x)} \mu^{n-1} \times \nu_x \left[\left(\bigcup_j A_j \times B_j\right) \cap (G \times I)\right].
\end{aligned}$$

Just as in the proof of Theorem 1 this gives that for all $U \in \mathcal{R}$ we have

$$\begin{aligned}
\mu_{n,x} \circ T^{-1}[U] - \mu^{n-1} \times \nu_x[U] &\ge -\left(\mu^{n-1} \times \nu_x\left[(G \times I)'\right] + \left|\frac{n\bar{F}(x)}{\bar{F}_n(x)} - 1\right|\right) \\
&\ge -\left(\mu^{n-1}[G'] + \nu_x[I'] + \left|\frac{n\bar{F}(x)}{\bar{F}_n(x)} - 1\right|\right) \\
&\ge -R(n, L, x) \quad \forall x \ge \ell_n.
\end{aligned}$$

In the previous equation and in the following the prime symbol denotes the complement of a set in the appropriate space: $(G \times I)' = \mathbb{R}^n \setminus (G \times I)$, $I' = \mathbb{R} \setminus I$ and $G' = \mathbb{R}^{n-1} \setminus G$.

Since $\mathcal{R}$ is closed under complementation we can also get an upper bound by applying the previous inequality for U' to get

$$|\mu_{n,x} \circ T^{-1}[U] - \mu^{n-1} \times \nu_x[U]| \le R(n, L, x).$$

The proof is now completed by letting x or $n \to \infty$, then $L \to \infty$ as before. □

Acknowledgements

The authors would like to thank Amir Dembo, Stefan Großkinsky, Richard Samworth, and an anonymous referee for their useful comments. IA has been partially supported by the PICT-2008-0315 project "Probability and Stochastic Processes". ML has been partially supported by the FP7-REGPOT-2009-1 project "Archimedes Center for Modeling, Analysis and Computation".

References

[1] Inés Armendáriz, Michail Loulakis, Thermodynamic limit for the invariant measures in supercritical zero range processes, Probab. Theory Related Fields 145 (1–2) (2009) 175–188.
[2] Søren Asmussen, Serguei Foss, Dmitry Korshunov, Asymptotics for sums of random variables with local subexponential behaviour, J. Theoret. Probab. 16 (2) (2003) 489–518.
[3] A.A. Balkema, L. de Haan, Residual life time at great age, Ann. Probab. 2 (1974) 792–804.
[4] N.H. Bingham, C.M. Goldie, J.L. Teugels, Regular Variation, in: Encyclopedia of Mathematics and its Applications, vol. 27, Cambridge University Press, Cambridge, 1989.
[5] D. Denisov, A.B. Dieker, V. Shneer, Large deviations for random walks under subexponentiality: the big-jump domain, Ann. Probab. 36 (5) (2008) 1946–1991.
[6] Richard Durrett, Conditioned limit theorems for random walks with negative drift, Z. Wahrsch. Verw. Gebiete 52 (3) (1980) 277–287.
[7] William Feller, An Introduction to Probability Theory and its Applications, second edition, vol. II, John Wiley & Sons Inc., New York, 1971.
[8] Pablo A. Ferrari, Claudio Landim, Valentin V. Sisko, Condensation for a fixed number of independent random variables, J. Stat. Phys. 128 (5) (2007) 1153–1158.
[9] Stefan Großkinsky, Gunter M. Schütz, Herbert Spohn, Condensation in the zero range process: stationary and dynamical properties, J. Stat. Phys. 113 (3-4) (2003) 389–410.
[10] C.C. Heyde, On large deviation probabilities in the case of attraction to a non-normal stable law, Sankhyā Ser. A 30 (1968) 253–258.
[11] T. Mikosch, A.V. Nagaev, Large deviations of heavy-tailed sums with applications in insurance, Extremes 1 (1) (1998) 81–110.
[12] A.V. Nagaev, Limit theorems that take into account large deviations when Cramér's condition is violated, Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk 13 (6) (1969) 17–22.
[13] S.V. Nagaev, Large deviations of sums of independent random variables, Ann. Probab. 7 (5) (1979) 745–789.
[14] Jozef L. Teugels, The class of subexponential distributions, Ann. Probab. 3 (6) (1975) 1000–1011.
[15] Vladimir Vinogradov, Refined Large Deviation Limit Theorems, in: Pitman Research Notes in Mathematics Series, vol. 315, Longman Scientific & Technical, Harlow, 1994.