



HAL
open science

A Gibbs Conditional theorem under extreme deviation

Maeva Biret, Michel Broniatowski, Zangsheng Cao

► **To cite this version:**

Maeva Biret, Michel Broniatowski, Zangsheng Cao. A Gibbs Conditional theorem under extreme deviation. 2016. hal-03897488v1

HAL Id: hal-03897488

<https://hal.science/hal-03897488v1>

Preprint submitted on 13 Oct 2016 (v1), last revised 14 Jan 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Gibbs Conditional theorem under extreme deviation

Maeva Biret⁽¹⁾, Michel Broniatowski^{(2,*),} Zangsheng Cao⁽²⁾
⁽¹⁾ SNECMA***, ⁽²⁾ LSTA, Université Pierre et Marie
Curie, Paris, France, ^(*) Corresponding author

Abstract

We explore some properties of the conditional distribution of an i.i.d. sample under large exceedances of its sum. Thresholds for the asymptotic independance of the summands are observed, in contrast with the classical case when the conditioning event is in the range of a large deviation. This paper is an extension to [7]. Tools include a new Edgeworth expansion adapted to specific triangular arrays where the rows are generated by tilted distribution with diverging parameters, together with some Abelian type results.

1 Introduction

Let $X_1^n := (X_1, \dots, X_n)$ be n independent unbounded real valued random variables and $S_1^n := X_1 + \dots + X_n$ denote their sum. The purpose of this paper is to explore the limit distribution of the generic variable X_1 conditioned on extreme deviations (ED) pertaining to S_1^n . By extreme deviation we mean that S_1^n/n is supposed to take values which are going to infinity as n increases. Obviously such events are of infinitesimal probability. Our interest in this question stems from a first result which assesses that under appropriate conditions, when the sequence a_n is such that

$$\lim_{n \rightarrow \infty} a_n = \infty$$

then there exists a sequence ε_n which tends to 0 as n tends to infinity such that

$$\lim_{n \rightarrow \infty} P(\cap_{i=1}^n (X_i \in (a_n - \varepsilon_n, a_n + \varepsilon_n)) | S_1^n/n > a_n) = 1, \quad (1)$$

which is to say that when the empirical mean takes exceedingly large values, then all the summands share the same behaviour. This result

obviously requires a number of hypotheses, which we simply quote as "light tails" type. We refer to [7] for this result and the connection with earlier related works; that such most unusual cases may be considered is argued in this latest paper, in relation with the Erdős-Rényi law of large numbers and the formation of high level aggregates in random sequences.

The above result is clearly to be put in relation with the so-called Gibbs conditional Principle which we recall briefly in its simplest form.

Consider the case when the sequence $a_n = a$ is constant with value larger than the expectation of X_1 . Hence we consider the behaviour of the summands when $(S_1^n/n > a)$, under a large deviation (LD) condition about the empirical mean. The asymptotic conditional distribution of X_1 given $(S_1^n/n > a)$ is the well known tilted distribution of P_X with parameter t associated to a . Let us introduce some notation to shed some light on this. The hypotheses to be stated now together with notation are kept throughout the entire paper. Without loss of generality it is assumed that the generic r.v. X_1 takes only non negative values.

It will be assumed that P_X , which is the distribution of X_1 , has a density p with respect to the Lebesgue measure on \mathbb{R} . The fact that X_1 has a light tail is captured in the hypothesis that X_1 has a moment generating function

$$\Phi(t) := E[\exp tX_1],$$

which is finite in a non void neighborhood \mathcal{N} of 0. This fact is usually referred to as a Cramer type condition.

Defined on \mathcal{N} are the following functions. The functions

$$t \rightarrow m(t) := \frac{d}{dt} \log \Phi(t) \tag{2}$$

$$t \rightarrow s^2(t) := \frac{d}{dt} m(t) \tag{3}$$

$$t \rightarrow \mu_j(t) := \frac{d^j}{dt^j} \log \Phi(t), \quad j \geq 3 \tag{4}$$

are the expectation, the variance, and the centered moments of order j of the r.v. \mathcal{X}_t with density

$$\pi_t(x) := \frac{\exp tx}{\Phi(t)} p(x)$$

which is defined on \mathbb{R} and which is the tilted density with parameter t . When Φ is steep, meaning that

$$\lim_{t \rightarrow t^+} m(t) = \infty$$

where $t^+ := \text{ess sup } \mathcal{N}$ then m parametrizes the convex hull of the support of P_X . We refer to Barndorff-Nielsen [3] for those properties. As a consequence of this fact, for all a in the support of P_X , it will be convenient to define

$$\pi^a = \pi_t$$

where a is the unique solution of the equation $m(t) = a$.

The Gibbs conditional principle in the standard above setting can be stated as follows.

As n tends to infinity the conditional distribution of X_1 given $(S_1^n/n > a)$ is Π^a , the distribution with density π^a .

Indeed we prefer to state Gibbs principle in a form where the conditioning event is a point condition $(S_1^n/n = a)$. The conditional distribution of X_1 given $(S_1^n/n = a)$ is a well defined distribution and Gibbs conditional principle states that this conditional distribution converges to Π^a as n tends to infinity. In both settings, this convergence holds in total variation norm. We refer to [11] for the local form of the conditioning event; we will mostly be interested in the extension of this form in the present paper.

For all α (depending on n or not) we will denote p_α the density of the random vector X_1^k conditioned upon the local event $(S_1^n = n\alpha)$. The notation $p_\alpha(X_1^k = x_1^k)$ is sometimes used to denote the value of the density p_α at point x_1^k . The same notation is used when X_1, \dots, X_k are sampled under some Π^α , namely $\pi^\alpha(X_1^k = x_1^k)$.

This article is organized as follows. Notation and hypotheses are stated in Section 2, along with some necessary facts from asymptotic analysis in the context of light tailed densities. Section 4 provides a local Gibbs conditional principle under EDP, namely producing the approximation of the conditional density of X_1 conditionally on $(S_1^n/n = a_n)$ for sequences a_n which tend to infinity. We explore two rates of growth for the sequence a_n , which yield two different approximating distributions for the conditional law of X_1 . The first one extends the classical approximation by the tilted one, substituting π^a by π^{a_n} . The second case, which corresponds to a faster growth of a_n , produces an approximation of a different kind. It may be possible to explore faster growth conditions than those considered here, leading to a wide class of approximating distributions; this would require some high order Edgeworth expansions for triangular arrays of variables, extending the corresponding result of order 3 presented in this paper; we did not move further in this direction, in order to avoid too many technicalities.

For fixed k and fixed $a_n = a > E(X_1)$ it is known that the r.v.'s X_1, \dots, X_k are asymptotically independent given $(S_1^n/n = a_n)$; see [11].

This statement is explored when a_n grows to infinity with n , keeping k fixed. It is shown that the asymptotic independence property holds for sequences a_n with moderate growth, and that independence fails for sequences a_n with fast growth.

The local approximation of the density of X_1 conditionally on $(S_1^n/n = a_n)$ is further extended to typical paths under the conditional sampling scheme, which in turn provides the approximation in variation norm for the conditional distribution; the method used here follows closely the approach by [8]. The differences between the Gibbs principles in LDP and EDP are discussed. Section 5 states similar results in the case when the conditioning event is $(S_1^n/n > a_n)$.

The main tools to be used come from asymptotic analysis and local limit theorems, developed from [13] and [4]; we also have borrowed a number of arguments from [16]. An Edgeworth expansion for some special array of independent r.v.'s with tilted distribution and argument moving to infinity with the row-size is needed; its proof is deferred to the Section 6. The basic Abelian type result which is used is stated in [6].

2 Notation and hypotheses

Thereafter we will use indifferently the notation $f(t) \underset{t \rightarrow \infty}{\sim} g(t)$ and $f(t) \underset{t \rightarrow \infty}{=} g(t)(1 + o(1))$ to specify that f and g are asymptotically equivalent functions.

The density p is assumed to be of the form

$$p(x) = \exp(-(g(x) - q(x))), \quad x \in \mathbb{R}_+. \quad (5)$$

The function q is assumed to be bounded, so that the asymptotic behaviour of p is captured through the function g . The function g is positive, convex, four times differentiable and satisfies

$$\frac{g(x)}{x} \underset{x \rightarrow \infty}{\longrightarrow} \infty. \quad (6)$$

Define

$$h(x) := g'(x). \quad (7)$$

In the present context, due to (6) and the assumed conditions on q to be stated hereunder, $t^+ = +\infty$.

Not all positive convex g 's satisfying (6) are adapted to our purpose. We follow the line of Juszczak and Nagaev [16] to describe the assumed regularity conditions of h . See also [1] for somehow similar conditions.

We firstly assume that the function h , which is a positive function defined on \mathbb{R}_+ , is either regularly or rapidly varying in a neighborhood

of infinity; the function h is monotone and, by (6), $h(x) \rightarrow \infty$ when $x \rightarrow \infty$.

The following notation is adopted.

$RV(\alpha)$ designates the class of regularly varying functions of index α defined on \mathbb{R}_+ ,

$$\psi(t) := h^{\leftarrow}(t)$$

designates the inverse of h . Hence ψ is monotone for large t and $\psi(t) \rightarrow \infty$ when $t \rightarrow \infty$, $\sigma^2(x) := 1/h'(x)$, $\hat{x} := \hat{x}(t) = \psi(t)$, $\hat{\sigma} := \sigma(\hat{x}) = \sigma(\psi(t))$.

The two cases considered for h , the regularly varying case and the rapidly varying case, are described below. The first one is adapted to regularly varying functions g , whose smoothness is described through the following condition pertaining to h .

The Regularly varying case. It will be assumed that h belongs to the subclass of $RV(\beta)$, $\beta > 0$, with

$$h(x) = x^\beta l(x),$$

where the Karamata form of the slowly varying function l takes the form

$$l(x) = c \exp \int_1^x \frac{\epsilon(u)}{u} du \quad (8)$$

for some positive c . We assume that $x \mapsto \epsilon(x)$ is twice differentiable and satisfies

$$\begin{cases} \epsilon(x) \underset{x \rightarrow \infty}{=} o(1), \\ x|\epsilon'(x)| \underset{x \rightarrow \infty}{=} O(1), \\ x^2|\epsilon^{(2)}(x)| \underset{x \rightarrow \infty}{=} O(1). \end{cases} \quad (9)$$

It will also be assumed that

$$|h^{(2)}(x)| \in RV(\theta) \quad (10)$$

where θ is a real number such that $\theta \leq \beta - 2$.

Remark 1 Under (8), when $\beta \neq 1$ then, under (10), $\theta = \beta - 2$. Whereas, when $\beta = 1$ then $\theta \leq \beta - 2$. A sufficient condition for the last assumption (10) is that $\epsilon'(t) \in RV(\gamma)$, for some $\gamma < -1$. Also in this case when $\beta = 1$, then $\theta = \beta + \gamma - 1$.

Example 2 Weibull density. Let p be a Weibull density with shape parameter $k > 1$ and scale parameter 1, namely

$$\begin{aligned} p(x) &= kx^{k-1} \exp(-x^k), \quad x \geq 0 \\ &= k \exp(-(x^k - (k-1) \log x)). \end{aligned}$$

Take $g(x) = x^k - (k-1)\log x$ and $q(x) = 0$. Then it holds

$$h(x) = kx^{k-1} - \frac{k-1}{x} = x^{k-1} \left(k - \frac{k-1}{x^k} \right).$$

Set $l(x) = k - (k-1)/x^k$, $x \geq 1$, which verifies

$$l'(x) = \frac{k(k-1)}{x^{k+1}} = \frac{l(x)\epsilon(x)}{x}$$

with

$$\epsilon(x) = \frac{k(k-1)}{kx^k - (k-1)}.$$

Since the function $\epsilon(x)$ satisfies the three conditions in (9), then $h(x) \in RV(k-1)$.

The Rapidly varying case. Here we have $h^{\leftarrow}(t) = \psi(t) \in RV(0)$ and

$$\psi(t) = c \exp \int_1^t \frac{\epsilon(u)}{u} du, \quad (11)$$

for some positive c , and $t \mapsto \epsilon(t)$ is twice differentiable with

$$\begin{cases} \epsilon(t) \underset{t \rightarrow \infty}{=} o(1), \\ \frac{t\epsilon'(t)}{\epsilon(t)} \underset{t \rightarrow \infty}{\rightarrow} 0, \\ \frac{t^2\epsilon^{(2)}(t)}{\epsilon(t)} \underset{t \rightarrow \infty}{\rightarrow} 0. \end{cases} \quad (12)$$

Note that these assumptions imply that $\epsilon(t) \in RV(0)$.

Example 3 A rapidly varying density. Define p through

$$p(x) = c \exp(-e^{x-1}), x \geq 0.$$

Then $g(x) = h(x) = e^{x-1}$ and $q(x) = 0$ for all non negative x . We show that $h(x)$ is a rapidly varying function. It holds $\psi(t) = \log t + 1$. Since $\psi'(t) = 1/t$, let $\epsilon(t) = 1/(\log t + 1)$ so that $\psi'(t) = \psi(t)\epsilon(t)/t$. Moreover, the three conditions of (12) are satisfied. Thus $\psi(t) \in RV(0)$ and $h(x)$ is a rapidly varying function.

Denote by \mathcal{R} the class of functions with either regular variation defined as in Case 2 or with rapid variation defined as in Case 2.

We now state hypotheses pertaining to the bounded function q in (5). We assume that

$$|q(x)| \in RV(\eta), \text{ for some } \eta < \theta - \frac{3\beta}{2} - \frac{3}{2} \text{ if } h \in RV(\beta) \quad (13)$$

and

$$|q(\psi(t))| \in RV(\eta), \text{ for some } \eta < -\frac{1}{2} \text{ if } h \text{ is rapidly varying.} \quad (14)$$

We will make use of the following result (see [6] Thm 3.1).

Theorem 4 Let $p(x)$ be defined as in (5) and $h(x)$ belong to \mathcal{R} . Denote by $m(t)$, $s^2(t)$ and $\mu_j(t)$ for $j = 3, 4, \dots$ the functions defined in (2), (3) and (4). Then it holds

$$\begin{aligned} m(t) &\underset{t \rightarrow \infty}{=} \psi(t)(1 + o(1)), \\ s^2(t) &\underset{t \rightarrow \infty}{=} \psi'(t)(1 + o(1)), \\ \mu_3(t) &\underset{t \rightarrow \infty}{=} \psi^{(2)}(t)(1 + o(1)), \\ \mu_j(t) &\underset{t \rightarrow \infty}{=} \begin{cases} M_j s^j(t)(1 + o(1)), & \text{for even } j > 3 \\ \frac{(M_{j+3} - 3jM_{j-1})\mu_3(t)s^{j-3}(t)}{6}(1 + o(1)), & \text{for odd } j > 3 \end{cases} \end{aligned}$$

where M_i , $i > 0$, denotes the i th order moment of standard normal distribution.

Corollary 5 Let $p(x)$ be defined as in (5) and $h(x) \in \mathfrak{R}$. Then it holds as $t \rightarrow \infty$

$$\frac{\mu_3(t)}{s^3(t)} \longrightarrow 0.$$

Proof. In the regularly varying case this follows from Corollaries 1 and 2 in [6], and in the rapidly varying case from Corollary 3 and Lemma 3 in [6]. ■

Our results require an extension of the classical Edgeworth expansions to triangular arrays of row-wise independent and identically distributed random variables, where the expectation of the generic r.v. in the n -th row tends to infinity with n . This can be achieved under log-concavity of p , i.e. when the function q is the null function, or when p is nearly log-concave. This is the scope of the next Section.

3 Edgeworth expansion under extreme normalizing factors

With π^{a_n} defined through

$$\pi^{a_n}(x) = \frac{e^{tx} p(x)}{\Phi(t)},$$

and t determined by $m(t) = a_n$ together with $s^2 := s^2(t)$ define the normalized density of π^{a_n} by

$$\bar{\pi}^{a_n}(x) = s\pi^{a_n}(sx + a_n),$$

and denote the n -convolution of $\bar{\pi}^{a_n}(x)$ by $\bar{\pi}_n^{a_n}(x)$. Denote by ρ_n its normalized density

$$\rho_n(x) := \sqrt{n}\bar{\pi}_n^{a_n}(\sqrt{n}x).$$

The following result extends the local Edgeworth expansion of the distribution of normalized sums of i.i.d. r.v.'s to the present context, where the summands are generated under the density $\bar{\pi}^{a_n}$. Therefore the setting is that of a triangular array of rowwise independent summands; the fact that $a_n \rightarrow \infty$ makes the situation unusual. We mainly adapt Feller's proof (Chapter 16, Theorem 2 [13]). However this variation on the classical Edgeworth expansion result requires some additional regularity assumption, which meet the requirements of Theorem 4, which are fulfilled in most models dealing with extremes and convolutions. Those are captured in cases when the density p is log-concave, or nearly log concave in the upper tail. Similar conditions are considered in [7].

Theorem 6 *With the above notation, uniformly upon x it holds*

$$\rho_n(x) = \phi(x) \left(1 + \frac{\mu_3}{6\sqrt{ns^3}}(x^3 - 3x) \right) + o\left(\frac{1}{\sqrt{n}}\right).$$

where $\phi(x)$ is standard normal density.

The proof of this result is postponed to the Section 6.

4 Gibbs' conditional principles under extreme events

We now explore Gibbs conditional principles under extreme events. The first result is a pointwise approximation of the conditional density $p_{a_n}(y_1)$ on \mathbb{R} . Two cases will be considered according to the rate of growth of the sequence a_n to infinity. For "moderate" growth our result extends the classical one pertaining to constant a_n larger than $E(X_1)$, since the approximating density of p_{a_n} is the tilted distribution with parameter a_n . For sequences a_n with fast growth, the approximating density includes a second order term which contributes to the approximation in a similar role as the tilted term; this term also appears in the first case, but is negligible with respect to the tilted density.

However this local approximation can be greatly improved when comparing p_{a_n} to its approximation. We will first prove that the approximation holds when the fixed arbitrary y_1 is substituted by a r.v. Y_1 with distribution P_{a_n} , henceforth on a typical realization under the distribution to be approximated. The approximation therefore holds in probability under this sampling scheme; a simple Lemma then proves that such a statement implies that the total variation distance between P_{a_n} and its approximation tends to 0 as n tends to infinity.

As a by-product we also address similar approximations for the case when the conditioning event writes $(S_1^n/n > a_n)$. The case when a_n grows to infinitely fast enough overlaps with that for which (1) holds.

Extension to the approximation of the distribution of X_1 given $(T_n = a_n)$ or $(T_n > a_n)$ where

$$T_n := \frac{1}{n} \sum_{i=1}^n f(X_i)$$

for functions f satisfying appropriate conditions are considered.

For sake of completeness we also provide some information when the density p_{a_n} is that of the vector (X_1, \dots, X_k) for fixed k . We prove that for moderate growth of a_n the approximating density is the product of corresponding marginal approximations, generalizing the well known result by Csiszar [9] which, in the present context, assesses the limit conditional independence of the coordinates of the vector (X_1, \dots, X_k) given $(S_n > na_n)$ for fixed $a_n > E(X_1)$ and fixed k . At the contrary this property is lost when a_n grows quickly to infinity.

Because of the property (1) it would be of interest to consider the joint distribution of the vector (X_1, \dots, X_{k_n}) given $(S_n > na_n)$ for sequences k_n close to n , as done in [8] for sequences a_n ranging from CLT to LDP. The extreme deviation case adds noticeable analytical difficulties.

4.1 A local result

Fix $y_1^k := (y_1, \dots, y_k)$ in \mathbb{R}^k and define $s_i^j := y_i + \dots + y_j$ for $1 \leq i < j \leq k$. Define t through

$$m := m(t) := a_n \tag{15}$$

and set

$$s := s(t)$$

for brevity.

We consider two conditions pertaining to the growth of the sequence a_n to infinity. In the first case we assume that

$$\lim_{n \rightarrow \infty} \frac{a_n}{s\sqrt{n}} = 0, \tag{16}$$

and in the second case we consider sequences a_n which may grow faster to infinity, obeying

$$0 < \liminf_{n \rightarrow \infty} \frac{a_n}{s\sqrt{n}} \leq \limsup_{n \rightarrow \infty} \frac{a_n}{s\sqrt{n}} < \infty. \tag{17}$$

Remark 7 *Both conditions (16) and (17) can be expressed in terms of a_n when the variance function $V(x)$ of the distribution of X_1 is known either in closed form, or is asymptotically equivalent to some known function. Recall that the variance function is defined on $Im(X_1)$ through*

$$x \rightarrow V(x) = s^2 om^{-1}(x).$$

See e.g. [2] for a description of distribution functions with polynomial variance function and [15] for tail equivalence for the variance function in infinitely divisible distributions. In the Regularly varying case, i.e. when h belongs to the subclass of $RV(\beta)$, $\beta > 0$, then standard operations on smooth regularly varying functions yield $V(x) = x^{1-\beta} l(x)$ for some slowly varying function l ; see [4]; hence $s(t) = a_n^{(1-\beta)/2} l(a_n)$. Assuming that $V(x) \sim x^{2\rho}$ as $x \rightarrow \infty$, it follows that (16) writes $a_n = o(n^{1/(1+\rho)})$ whereas (17) amounts to assume that a_n is of order $n^{1/(1+\rho)}$.

Denote

$$z := \frac{m - y_1}{s\sqrt{n-1}}.$$

Theorem 8 *When (16) holds then it holds*

$$p_{a_n}(y_1) = p(X_1 = y_1 | S_1^n = na_n) = \pi^{a_n}(y_1) \left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right),$$

The proof of this result is postponed to the Section 6.

Remark 9 *The above condition (16) is not sufficient to entail (1) to hold. This yields to the study of a similar limit conditional result under the corresponding condition (17).*

Theorem 10 *Assume that the sequence a_n satisfies (17). Denote*

$$\alpha := t + \frac{\mu_3}{2(n-1)s^2}$$

and

$$\beta := (n-1)s^2.$$

Then

$$p(X_1 = y_1 | S_1^n = na_n) = g_{a_n}(y_1) := Cp(y_1) \mathbf{n}(\alpha\beta + a_n, \beta, y_1) (1 + o(1)) \quad (18)$$

where $\mathbf{n}(\mu, \sigma^2, x)$ denotes the normal density with expectation μ and variance σ^2 evaluated at point x , and C is a normalizing constant.

When (16) holds instead of (17) then

$$p(X_1 = y_1 | S_1^n = na_n) = \pi^{a_n}(y_1)(1 + o(1))$$

for all y_1 as n tends to infinity.

Proof. In contrast with the above case, the second summand in (39) does not tend to 0 any longer and contributes to the approximating density. Standard development then yields the result. When (16) holds instead of (17) then standard expansions in (18) provide $g_{a_n}(y_1) \sim \pi^{a_n}(y_1)$ for all y_1 as n tends to infinity. ■

4.2 On conditional independence under extreme events

We now turn to the case when we approximate the joint conditional density $p_{a_n}(y_1^k) := p_{a_n}(y_1, \dots, y_k) = p_{a_n}(X_1^k = y_1^k | S_1^n = na_n)$. Denote $s_i^j := y_i + \dots + y_j$ for $i \leq j$ and $s_1^0 := 0$.

We first consider the case when (16) holds. We then have

Proposition 11 *When (16) holds then for any fixed k*

$$p_{a_n}(y_1^k) = \prod_{i=1}^k \pi^{m_i}(y_i) (1 + o(1/\sqrt{n}))$$

where

$$m_i := m(t_i) := \frac{na_n - s_1^i}{n - i}. \quad (19)$$

The proof of this result is postponed to the Section 6.

We now explore the limit conditional independence of blocks of fixed length under extreme condition. As a consequence of the above Proposition 11 it holds

Theorem 12 *Under (16) it holds*

$$p_{a_n}(y_1^k) = p(X_1^k = y_1^k | S_1^n = na_n) = \left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right) \prod_{i=1}^k \pi^{a_n}(X_i = y_i),$$

The technical proof is deferred to the Section 6.

Remark 13 *The above result shows that asymptotically the point condition ($S_1^n = na_n$) leaves blocks of k of the X_i 's independent. Obviously this property does not hold for large values of k , close to n . A similar statement holds in the LDP range, conditioning either on ($S_1^n = na$) (see Diaconis and Friedman [11]), or on ($S_1^n \geq na$) (see Csiszar [9] for a general statement on asymptotic conditional independence given events with positive probability).*

We now turn to the case when a_n moves more quickly to infinity. Denote

$$m_i := m(t_i) := \frac{na_n - s_1^{i-1}}{n - i + 1}$$

together with

$$s_i^2 := s^2(t_i).$$

Theorem 14 *Assume that (17) holds. Then for all fixed k it holds*

$$p(X_1^k = y_1^k | S_1^n = na_n) = \prod_{i=1}^k g_i(y_i) (1 + o(1))$$

where

$$g_i(y_i) := C_i p(y_i) \mathbf{n}(\alpha_i \beta_i + a_n, \beta_i, y_i)$$

and

$$\alpha_i := t_i + \frac{\mu_3}{2(n-i+1)s_i^2} \quad (20)$$

$$\beta_i := (n-i+1)s_i^2. \quad (21)$$

Remark 15 *When (16) holds, the above result is a refinement of the result in Proposition 11. Under (17) and when (16) does not hold, the approximations obtained in Lemma 31 do not hold, and the approximating density cannot be stated as a product of densities under which independence holds. In that case it follows that the conditional independence property under extreme events does not hold any longer.*

4.3 Strengthening the local Gibbs conditional principle

We now turn to a stronger approximation of p_{a_n} . Consider Y_1 a r.v. with density p_{a_n} , and the random variable $p_{a_n}(Y_1) := p(X_1 = Y_1 | S_1^n = na_n)$. Denote

$$g_{a_n}(x) := Cp(x) \mathbf{n}(\alpha\beta + a_n, \beta, x)$$

where $\alpha := \alpha_n$ and $\beta := \beta_n$ are defined in (20) and (21), and C is a normalizing constant, it holds

Theorem 16 (i) *When (16) holds then*

$$p_{a_n}(Y_1) = \pi^{a_n}(Y_1) (1 + R_n)$$

where the tilted density at point a_n , and where R_n is a function of Y_1 such that $P_{a_n}(|R_n| > \delta\sqrt{n}) \rightarrow 0$ as $n \rightarrow \infty$ for any positive δ . When (17) holds then, with t_n such that $m(t_n) = a_n$, $\alpha := \alpha_n$ and $\beta := \beta_n$

$$p_{a_n}(Y_1) = g_{a_n}(Y_1)(1 + R'_n)$$

where $P_{a_n}(|R'_n| > \delta) \rightarrow 0$ as $n \rightarrow \infty$ for any positive δ .

Remark 17 *This result is of much greater relevance than the previous ones. Indeed under P_{a_n} the r.v. Y_1 may take large values. On the contrary simple approximation of p_{a_n} by π^{a_n} or g_{a_n} on \mathbb{R}_+ only provides some knowledge on p_{a_n} on sets with smaller and smaller probability under p_{a_n} . Also it will be proved that as a consequence of the above result, the L^1 norm between p_{a_n} and its approximation goes to 0 as $n \rightarrow \infty$, a result out of reach through the aforementioned results.*

In order to adapt the proof of Theorem 12 to the present setting it is necessary to get some insight on the plausible values of Y_1 under P_{a_n} . It holds

Lemma 18 *It holds*

$$Y_1 = O_{P_{a_n}}(a_n).$$

Proof. This is a consequence of Markov Inequality:

$$P(Y_1 > u | S_1^n = na_n) \leq \frac{E(Y_1 | S_1^n = na_n)}{u} = \frac{a_n}{u}$$

which goes to 0 for all $u = u_n$ such that $\lim_{n \rightarrow \infty} u_n/a_n = \infty$.

Now making use of Lemma 18 in the proof of Theorem 8 and Theorem 10, substituting y_1 with Y_1 , completes the proof. ■

Denote the probability measures P_{a_n} , Π^{a_n} and G_{a_n} with respective densities p_{a_n} , π^{a_n} and g_{a_n} .

4.4 Gibbs principle in variation norm

We now consider the approximation of P_{a_n} by G_{a_n} in variation norm.

The main ingredient is the fact that in the present setting approximation of p_{a_n} by g_{a_n} in probability plus some rate implies approximation of the corresponding measures in variation norm. This approach has been developed in [8]; we state a first lemma which states that whether two densities are equivalent in probability with small relative error when measured according to the first one, then the same holds under the sampling of the second.

Let \mathfrak{R}_n and \mathfrak{S}_n denote two p.m.'s on \mathbb{R}^n with respective densities \mathfrak{r}_n and \mathfrak{s}_n .

Lemma 19 *Suppose that for some sequence ϖ_n which tends to 0 as n tends to infinity*

$$\mathfrak{r}_n(Y_1^n) = \mathfrak{s}_n(Y_1^n) (1 + o_{\mathfrak{R}_n}(1)) \quad (22)$$

as n tends to ∞ . Then

$$\mathfrak{s}_n(Y_1^n) = \mathfrak{r}_n(Y_1^n) (1 + o_{\mathfrak{S}_n}(1)). \quad (23)$$

The proof of this result is available in [8]. Applying this Lemma to the present setting yields

$$g_{a_n}(Y_1) = p_{a_n}(Y_1) \left(1 + o_{G_{a_n}}(1/\sqrt{n})\right)$$

as $n \rightarrow \infty$, which together with Theorem 10 or Theorem 8 implies

$$p_{a_n}(Y_1) = g_{a_n}(Y_1) \left(1 + o_{P_{a_n}}(1/\sqrt{n})\right)$$

or

$$p_{a_n}(Y_1) = \pi^{a_n}(Y_1) \left(1 + o_{P_{a_n}}(1/\sqrt{n})\right)$$

This fact entails

Theorem 20 *Under (17) the total variation norm between P_{a_n} and G_{a_n} goes to 0 as $n \rightarrow \infty$. When (16) holds then the total variation norm between P_{a_n} and Π^{a_n} goes to 0 as $n \rightarrow \infty$.*

The proof of this theorem is also provided in [8].

Remark 21 *This result is to be paralleled with Theorem 1.6 in Diaconis and Freedman [11] and Theorem 2.15 in Dembo and Zeitouni [10] which provide a rate for this convergence in the LDP range.*

4.5 The asymptotic location of X under the conditioned distribution

This paragraph intends to provide some insight on the behaviour of X_1 under the condition ($S_1^n = na_n$); this will be extended further on to the case when ($S_1^n \geq na_n$) and to be considered in parallel with similar facts developed in [8] for larger values of a_n .

Let \mathcal{X}_t be a r.v. with density π^{a_n} where $m(t) = a_n$ and a_n satisfies (16) or (17). Recall that $E\mathcal{X}_t = a_n$ and $Var\mathcal{X}_t = s^2$. We evaluate the moment generating function of the normalized variable $(\mathcal{X}_t - a_n)/s$. It holds

$$\log E[\exp(\lambda (\mathcal{X}_t - a_n) / s)] = -\lambda a_n / s + \log \Phi \left(t + \frac{\lambda}{s} \right) - \log \Phi(t).$$

A second order Taylor expansion in the above display yields

$$\log E[\exp(\lambda (\mathcal{X}_t - a_n) / s)] = \frac{\lambda^2 s^2}{2} \frac{\left(t + \frac{\theta\lambda}{s}\right)}{s^2}$$

where $\theta = \theta(t, \lambda) \in (0, 1)$. The proof of the following Lemma is deferred to the Section 6. It holds

Lemma 22 *Under the above hypotheses and notation, for any compact set K ,*

$$\lim_{n \rightarrow \infty} \sup_{u \in K} \frac{s^2 \left(t + \frac{u}{s}\right)}{s^2} = 1.$$

Applying the above Lemma it follows that the normalized r.v.'s $(\mathcal{X}_t - a_n) / s$ converge to a standard normal variable $N(0, 1)$ in distribution, as $n \rightarrow \infty$. This amounts to say that

$$\mathcal{X}_t = a_n + sN(0, 1) + o_{\Pi^{a_n}}(1).$$

which implies that \mathcal{X}_t concentrates around a_n with rate s . Due to Theorem 20 the same holds for X_1 under $(S_1^n = na_n)$.

4.6 Conditional limit behaviour under other mean effect events

Let X_1, \dots, X_n denote n i.i.d. real valued r.v.'s with distribution P and density p and let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function such that $\Phi_f(\lambda) := E[\exp(\lambda f(X_1))]$ is finite for λ in a non void neighborhood of 0 (the so-called Cramer condition). Denote $m_f(\lambda)$ and $s_f^2(\lambda)$ the first and second derivatives of $\log \Phi_f(\lambda)$. Assume that the r.v. $f(X_1)$ has density p_f on \mathbb{R} , and denote $p_f(f(X_1) = u)$ its value at point u .

Denote

$$\pi_f^a(y) = \frac{\exp \lambda y}{\Phi_f(\lambda)} p_f(y)$$

with λ the unique solution of the equation $m_f(\lambda) = a$ for all a in $Im(f(X_1))$ assuming that $\lambda \rightarrow \Phi_f(\lambda)$ is steep on its domain. Denote

$$F_i^j := f(X_i) + \dots + f(X_j)$$

for $1 \leq i \leq j \leq n$. We make use of the following equality

$$\begin{aligned} p(X_1 = x | F_1^n = na_n) &= \frac{p(X_1 = x)}{p_f(f(X_1) = f(x))} \times \\ &\left(p_f(f(X_1) = f(x)) \frac{p_f(f(X_1) = f(x) | F_1^n = na_n)}{p_f(f(X_1) = f(x) | F_2^n = na_n - f(x))} \right). \end{aligned}$$

Note that for all α in $Im(f(X_1))$, denoting λ the solution of $m_f(\lambda) = \alpha$ and defining

$${}_f\pi^\alpha(x) := \frac{e^{\lambda f(x)} p(X = x)}{\int e^{\lambda f(x)} p(X = x) dx}$$

it is readily checked that

$${}_f\pi^\alpha(x) = \frac{p(X_1 = x)}{p_f(f(X_1) = f(x))} \pi_f^{a_n}(f(x)).$$

Denoting $P_{a_n, f}$ the distribution of X_1 given ($F_1^n = na_n$) it results, using Theorem 16 that the following Theorem holds.

Theorem 23 *Assume that, with s substituted by s_f , condition (16) holds. Then*

$$p(X_1 = Z | F_1^n = na_n) = {}_f\pi^\alpha(Z) (1 + o_{P_{a_n, f}}(1/\sqrt{n}))$$

and under (17)

$$p(X_1 = Z | F_1^n = na_n) = C \frac{p(X_1 = Z)}{p_f(f(X_1) = f(Z))} \mathbf{n}(\alpha\beta + a_n, \beta, f(Z)) (1 + o_{P_{a_n, f}}(1/\sqrt{n}))$$

where $\alpha := \alpha_n$ and $\beta := \beta_n$ are defined in (20) and (21) with m and s substituted by m_f and s_f .

Remark 24 *The first part of the above Theorem extends the classical Gibbs Principle under condition (16), which, for fixed $a = a_n$ writes*

$$p(X_1 = x | F_1^n = na) = {}_f\pi^\alpha(x) (1 + o(1/\sqrt{n}))$$

for any fixed x . See [11]. This statement does not hold any longer under condition (17).

Remark 25 *Making use of the same arguments as in Subsection 4.4 it follows that Theorem 23 yields that the variation distance between the conditional distribution and its approximation tends to 0 as n tends to infinity.*

Example 26 *Consider for example the application of the above result to r.v's Y_1, \dots, Y_n with $Y_i := (X_i)^2$ where the X_i 's are i.i.d. and are such that the density of the i.i.d. r.v's Y_i 's satisfy (5), where $h \in R_\beta \cup R_\infty$ with $\beta > 1$. By the Gibbs conditional principle, for fixed a , conditionally on $(\sum_{i=1}^n Y_i = na)$ the generic r.v. Y_1 has a non degenerate limit distribution*

$$p_Y^*(y) := \frac{\exp ty}{E \exp tY_1} p_Y(y)$$

and the limit density of X_1 under $(\sum_{i=1}^n X_i^2 = na)$ is

$$p_X^*(y) := \frac{\exp tx^2}{E \exp tX_1^2} p_X(y)$$

whereas, when $a_n \rightarrow \infty$, Y_1 's the limit conditional distribution is degenerate and concentrates around a_n . As a consequence the distribution of X_1 under the condition $(\sum_{i=1}^n X_i^2 = na_n)$ concentrates sharply at $-\sqrt{a_n}$ and $+\sqrt{a_n}$.

5 EDP under exceedance

The following proposition states the marginally conditional density under condition $A_n = \{S_1^n \geq na_n\}$. We denote this density by p_{A_n} to differentiate it from p_{a_n} which is under condition $\{S_1^n = na_n\}$. For the purpose of the proof, we need the following Lemma, based on Theorem 6.2.1 of Jensen [14] in order to provide the asymptotic estimation of the tail probability $P(S_1^n \geq na_n)$ and of the n -convolution density $p(S_1^n/n = u)$ for $u > a_n$.

Define

$$I(x) := xm^{-1}(x) - \log \Phi(m^{-1}(x)).$$

We make use of the following result (see Section 6 for the proof).

Lemma 27 *Set $m(t) = a_n$. Suppose that $a_n \rightarrow \infty$ as $n \rightarrow \infty$. Then it holds*

$$P(S_1^n \geq na_n) = \frac{\exp(-nI(a_n))}{\sqrt{2\pi}\sqrt{nts}(t)} \left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right). \quad (24)$$

Let further t_τ be defined by $m(t_\tau) = \tau$ with $\tau \geq a_n$, it then holds, uniformly upon τ

$$p(S_1^n = n\tau) = \frac{\sqrt{n} \exp(-nI(\tau))}{\sqrt{2\pi}s(t_\tau)} \left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right). \quad (25)$$

The proof of this lemma is postponed to the Section 6.

Theorem 28 *Let X_1, \dots, X_n be i.i.d. random variables with density $p(x)$ defined in (5) and $h(x) \in \mathcal{R}$. Set $m(t) = a_n$ let η_n be a positive sequence satisfying*

$$\eta_n \rightarrow 0 \quad \text{and} \quad nm^{-1}(a_n)\eta_n \rightarrow \infty.$$

(i) *When (16) holds*

$$p_{A_n}(y_1) = p(X_1 = y_1 | S_1^n \geq na_n) = \pi_{A_n}(y_1) \left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right),$$

with

$$\pi_{A_n}(y_1) = ts(t) e^{nI(a_n)} \int_{a_n}^{a_n + \eta_n} \pi_\tau(y_1) \exp(-nI(\tau) - \log s(t_\tau)) d\tau$$

with t_τ defined by $m(t_\tau) = \tau$.

(ii) *When (17) holds*

$$p_{A_n}(y_1) = p(X_1 = y_1 | S_1^n \geq na_n) = g_{A_n}(y_1) \left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right),$$

with

$$g_{A_n}(y_1) = ts(t)e^{nI(a_n)} \int_{a_n}^{a_n + \eta_n} g_\tau(y_1) \exp(-nI(\tau) - \log s(t_\tau)) d\tau,$$

where $g_\tau = \pi^\tau$ with t_τ defined by $m(t_\tau) = \tau$.

The proof is postponed to the Section 6.

Remark 29 Conditions (17) and (16) have to be compared with the growth condition pertaining to the sequence a_n for which (1) holds. Consider the regularly varying case, namely assume that $h(x) = x^\beta l(x)$ for some $\beta > 0$. Then making use of Theorem 4 it is readily checked that (17) amounts to

$$\liminf_{n \rightarrow \infty} \frac{a_n}{n^{1/(1+\beta)}} > 0. \quad (26)$$

Now (1) holds whether for some $\delta > 1/(\beta + 1)$

$$\liminf_{n \rightarrow \infty} \frac{a_n}{n^\delta} > 0. \quad (27)$$

Assume that (27) holds; then (26) holds for all distributions p with $h(x) = x^\beta l(x)$ and $\beta > (1 - \delta)/\delta$. This can be stated as follows: Assume that for some $0 < \eta < 1$ it holds

$$\liminf_{n \rightarrow \infty} \frac{a_n}{n^\eta} > 0$$

then whenever $\beta > (1 - \eta)/\eta$ (27) and (26) simultaneously hold.

6 Appendix

6.1 Proof of Theorem 6

We state a preliminary Lemma, whose role is to provide some information on the characteristic function of the normalised random variable $(\mathcal{X}_t - m(t))/s(t)$ with density $\tilde{\pi}_t$ defined by

$$\tilde{\pi}_t(x) := \frac{s(t) \exp t(s(t)x + m(t)) p(s(t)x + m(t))}{\phi(t)} \quad (28)$$

as $t \rightarrow \infty$. The density p satisfies the hypotheses in Section 2. Denote $\varphi^{a_n}(u) := \int e^{iux} \tilde{\pi}_t(x) dx$ the characteristic function of $(\mathcal{X}_t - m(t))/s(t)$. It holds

Lemma 30 Assume that there exists c_1, c_2 both positive such that for all t

$$\tilde{\pi}_t(x) > c_1 \text{ for } |x| < c_2 \quad (29)$$

then under the hypotheses stated in Section 2, for any $c > 0$ there exists $\rho < 1$ such that

$$|\varphi^{a_n}(u)| \leq \rho \quad (30)$$

for $|u| > c$ and all a_n .

Proof. The proof of this Lemma is in [14], p150; we state it for completeness. Assume (29) holds with $\tilde{\pi}_t(x) > c_1$ for $|x| > c_2$ and setting $\epsilon := c_2/2$

$$\begin{aligned} |\varphi^{a_n}(u)| &\leq \left| \int e^{izu} 1(|z| < \epsilon) c_1 dz \right| + \int \{ \tilde{\pi}_t(z) - 1(|z| < \epsilon) c_1 \} dz \\ &\leq c_1 (2\epsilon) \left| \frac{e^{iu\epsilon} - e^{-iu\epsilon}}{2iu\epsilon} \right| + \{1 - 2\epsilon c_1\} \end{aligned}$$

and the last expression is independent on a_n and is such that for any $c > 0$ there exists $\rho < 1$ such that the expression is less than ρ for $|u| > c$.

For the density function $p(x)$ Theorem 5.4 of Nagaev [16] states that the normalized tilted density of $p(x)$, namely, $\tilde{\pi}_t(x)$ has the property

$$\lim_{a_n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\tilde{\pi}_t(x) - \varphi(x)| = 0 \quad (31)$$

which proves (29). ■

We now turn to the Proof of Theorem 6.

Since the proof is based on characteristic function (c.f.) arguments, we will use the following notation, in accordance with the common use in this area, therefore turning from laplace transform notation to characteristic function ones. Recall that we denote $\tilde{\pi}_t$ the normalized conjugate density of $p(x)$. Also ρ_n is the normalized n -fold convolution of $\tilde{\pi}_t$. Hence we consider the triangular array whose n -th row consists in n i.i.d. copies of a r.v. with standardized density $\tilde{\pi}_t$ and the sum of the row, divided by \sqrt{n} , has density ρ_n . The standard Gaussian density is denoted ϕ . The c.f. of $\tilde{\pi}_t$ is denoted φ^{a_n} so that the c.f. of ρ_n is $(\varphi^{a_n}(\cdot))^n$, and $m(t) = a_n$.

Step 1: In this step, we will express the following formula $G(x)$ by its Fourier transform. Let

$$G(x) := \rho_n(x) - \phi(x) - \frac{\mu_3}{6\sqrt{n}s_n^3} (x^3 - 3x) \phi(x).$$

From

$$\phi(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau x} e^{-\frac{1}{2}\tau^2} d\tau,$$

it follows that

$$\phi'''(x) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} (i\tau)^3 e^{-i\tau x} e^{-\frac{1}{2}\tau^2} d\tau.$$

On the other hand

$$\phi'''(x) = -(x^3 - 3x)\phi(x),$$

which gives

$$(x^3 - 3x)\phi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (i\tau)^3 e^{-i\tau x} e^{-\frac{1}{2}\tau^2} d\tau. \quad (32)$$

By Fourier inversion

$$\rho_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau x} (\varphi^{a_n}(\tau/\sqrt{n}))^n d\tau. \quad (33)$$

Using (32) and (33), we have

$$G(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\tau x} \left(\varphi^{a_n}(\tau/\sqrt{n})^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3} (i\tau)^3 e^{-\frac{1}{2}\tau^2} \right) d\tau.$$

Hence it holds

$$\begin{aligned} & \left| \rho_n(x) - \phi(x) - \frac{\mu_3}{6\sqrt{n}s^3} (x^3 - 3x)\phi(x) \right| \\ & \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| (\varphi^{a_n}(\tau/\sqrt{n}))^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3} (i\tau)^3 e^{-\frac{1}{2}\tau^2} \right| d\tau. \end{aligned}$$

Step 2: In this step, we show that for large n , the characteristic function φ^{a_n} satisfies

$$\int |\varphi^{a_n}(\tau)|^2 d\tau < \infty$$

By Parseval identity

$$\int |\varphi^{a_n}(\tau)|^2 d\tau = 2\pi \int (\tilde{\pi}_t(x))^2 dx \leq 2\pi \sup_{x \in \mathbb{R}} \tilde{\pi}_t(x) < \infty.$$

Use (31) to conclude the proof.

Step 3: In this step, we complete the proof by showing that when $n \rightarrow \infty$

$$\int_{-\infty}^{\infty} \left| (\varphi^{a_n}(\tau/\sqrt{n}))^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3} (i\tau)^3 e^{-\frac{1}{2}\tau^2} \right| d\tau = o\left(\frac{1}{\sqrt{n}}\right). \quad (34)$$

The LHS in (34) is splitted on $|\tau| > \omega\sqrt{n}$ and on $|\tau| \leq \omega\sqrt{n}$. It holds

$$\begin{aligned}
& \sqrt{n} \int_{|\tau| > \omega\sqrt{n}} \left| (\varphi^{a_n}(\tau/\sqrt{n}))^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3} (i\tau)^3 e^{-\frac{1}{2}\tau^2} \right| d\tau \\
& \leq \sqrt{n} \int_{|\tau| > \omega\sqrt{n}} \left| (\varphi^{a_n}(\tau/\sqrt{n})) \right|^n d\tau + \sqrt{n} \int_{|\tau| > \omega\sqrt{n}} \left| e^{-\frac{1}{2}\tau^2} + \frac{\mu_3}{6\sqrt{n}s^3} (i\tau)^3 e^{-\frac{1}{2}\tau^2} \right| d\tau \\
& \leq \sqrt{n}\rho^{n-2} \int_{|\tau| > \omega\sqrt{n}} \left| (\varphi^{a_n}(\tau/\sqrt{n})) \right|^2 d\tau + \sqrt{n} \int_{|\tau| > \omega\sqrt{n}} e^{-\frac{1}{2}\tau^2} \left(1 + \left| \frac{\mu_3\tau^3}{6\sqrt{n}s^3} \right| \right) d\tau.
\end{aligned} \tag{35}$$

where we used Lemma 30 from the second line to the third one. The first term of the last line tends to 0 when $n \rightarrow \infty$, since

$$\begin{aligned}
& \sqrt{n}\rho^{n-2} \int_{|\tau| > \omega\sqrt{n}} \left| (\varphi^{a_n}(\tau/\sqrt{n})) \right|^2 d\tau \\
& = \exp \left(\frac{1}{2} \log n + (n-2) \log \rho + \log \int_{|\tau| > \omega\sqrt{n}} (\varphi^{a_n}(\tau/\sqrt{n}))^2 d\tau \right) \rightarrow 0.
\end{aligned}$$

By Corollary 5 when $n \rightarrow \infty$

$$\begin{aligned}
& \sqrt{n} \int_{|\tau| > \omega\sqrt{n}} e^{-\frac{1}{2}\tau^2} \left(1 + \left| \frac{\mu_3\tau^3}{6\sqrt{n}s^3} \right| \right) d\tau \\
& \leq \sqrt{n} \int_{|\tau| > \omega\sqrt{n}} e^{-\frac{1}{2}\tau^2} |\tau|^3 d\tau = \sqrt{n} \int_{|\tau| > \omega\sqrt{n}} \exp \left\{ -\frac{1}{2}\tau^2 + 3 \log |\tau| \right\} d\tau \\
& = 2\sqrt{n} \exp \left(-\omega^2 n/2 + o(\omega^2 n/2) \right) \rightarrow 0,
\end{aligned}$$

where the second equality holds from, for example, Chapter 4 of [4]. Summing up, when $n \rightarrow \infty$

$$\int_{|\tau| > \omega\sqrt{n}} \left| (\varphi^{a_n}(\tau/\sqrt{n}))^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3} (i\tau)^3 e^{-\frac{1}{2}\tau^2} \right| d\tau = o\left(\frac{1}{\sqrt{n}}\right).$$

If $|\tau| \leq \omega\sqrt{n}$, it holds

$$\begin{aligned}
& \int_{|\tau| \leq \omega\sqrt{n}} \left| (\varphi^{a_n}(\tau/\sqrt{n}))^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3} (i\tau)^3 e^{-\frac{1}{2}\tau^2} \right| d\tau \\
& = \int_{|\tau| \leq \omega\sqrt{n}} e^{-\frac{1}{2}\tau^2} \left| (\varphi^{a_n}(\tau/\sqrt{n}))^n e^{\frac{1}{2}\tau^2} - 1 - \frac{\mu_3}{6\sqrt{n}s^3} (i\tau)^3 \right| d\tau \\
& = \int_{|\tau| \leq \omega\sqrt{n}} e^{-\frac{1}{2}\tau^2} \left| \exp \left\{ n \log \varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2 \right\} - 1 - \frac{\mu_3}{6\sqrt{n}s^3} (i\tau)^3 \right| d\tau.
\end{aligned} \tag{36}$$

The integrand in the last display is bounded through

$$|e^\alpha - 1 - \beta| = |(e^\alpha - e^\beta) + (e^\beta - 1 - \beta)| \leq (|\alpha - \beta| + \frac{1}{2}\beta^2)e^\gamma,$$

where $\gamma \geq \max(|\alpha|, |\beta|)$; this inequality follows replacing e^α, e^β by their power series, for real or complex α, β . Denote by

$$\gamma(\tau) = \log \varphi^{a_n}(\tau) + \frac{1}{2}\tau^2.$$

Since $\gamma'(0) = \gamma''(0) = 0$, the third order Taylor expansion of $\gamma(\tau)$ at $\tau = 0$ yields

$$\gamma(\tau) = \gamma(0) + \gamma'(0)\tau + \frac{1}{2}\gamma''(0)\tau^2 + \frac{1}{6}\gamma'''(\xi)\tau^3 = \frac{1}{6}\gamma'''(\xi)\tau^3,$$

where $0 < \xi < \tau$. Hence it holds

$$\left| \gamma(\tau) - \frac{\mu_3}{6s^3}(i\tau)^3 \right| = \left| \gamma'''(\xi) - \frac{\mu_3}{s_n^3}i^3 \right| \frac{\tau^3}{6}.$$

Here γ''' is continuous; thus we can choose ω small enough such that $|\gamma'''(\xi)| < \rho$ for $|\tau| < \omega$. Meanwhile, for n large enough, according to Corollary 5, we have $\mu_3/s^3 \rightarrow 0$. Hence it holds for n large enough

$$\left| \gamma(\tau) - \frac{\mu_3}{6s^3}(i\tau)^3 \right| \leq \left(|\gamma'''(\xi)| + \rho \right) \frac{|\tau|^3}{6} < \rho\tau^3. \quad (37)$$

Choose ω small enough, such that for n large enough it holds for $|\tau| < \omega$

$$\left| \frac{\mu_3}{6s^3}(i\tau)^3 \right| \leq \frac{1}{4}\tau^2, \text{ and } |\gamma(\tau)| \leq \frac{1}{4}\tau^2.$$

For this choice of ω , when $|\tau| < \omega$ we have

$$\max \left(\left| \frac{\mu_3}{6s^3}(i\tau)^3 \right|, |\gamma(\tau)| \right) \leq \frac{1}{4}\tau^2.$$

Replacing τ by τ/\sqrt{n} , it holds for $|\tau| < \omega\sqrt{n}$, and using (37)

$$\begin{aligned} & \left| n \log \varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2 - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 \right| \\ &= n \left| \log \varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\left(\frac{\tau}{\sqrt{n}}\right)^2 - \frac{\mu_3}{6s^3}\left(\frac{i\tau}{\sqrt{n}}\right)^3 \right| \\ &= n \left| \gamma\left(\frac{\tau}{\sqrt{n}}\right) - \frac{\mu_3}{6s^3}\left(\frac{i\tau}{\sqrt{n}}\right)^3 \right| < \frac{\rho|\tau|^3}{\sqrt{n}}. \end{aligned}$$

In a similar way, it also holds for $|\tau| < \omega\sqrt{n}$

$$\begin{aligned} & \max \left(\left| n \log \varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2 \right|, \left| \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 \right| \right) \\ &= n \max \left(\left| \gamma\left(\frac{\tau}{\sqrt{n}}\right) \right|, \left| \frac{\mu_3}{6s^3}\left(\frac{i\tau}{\sqrt{n}}\right)^3 \right| \right) \leq \frac{1}{4}\tau^2. \end{aligned}$$

Turn to the integrand in (36). We then for $|\tau| < \omega\sqrt{n}$

$$\begin{aligned} & \left| \exp \left\{ n \log \varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2 \right\} - 1 - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 \right| \\ & \leq \left(\left| n \log \varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2 - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 \right| + \frac{1}{2} \left| \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 \right|^2 \right) \\ & \quad \times \exp \left[\max \left(\left| n \log \varphi^{a_n}(\tau/\sqrt{n}) + \frac{1}{2}\tau^2 \right|, \left| \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 \right| \right) \right] \\ & \leq \left(\frac{\rho|\tau|^3}{\sqrt{n}} + \frac{1}{2} \left| \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 \right|^2 \right) \exp \left(\frac{\tau^2}{4} \right) \\ & = \left(\frac{\rho|\tau|^3}{\sqrt{n}} + \frac{\mu_3^2\tau^6}{72ns^6} \right) \exp \left(\frac{\tau^2}{4} \right). \end{aligned}$$

Use this upper bound to obtain

$$\begin{aligned} & \int_{|\tau| \leq \omega\sqrt{n}} \left| (\varphi^{a_n}(\tau/\sqrt{n}))^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 e^{-\frac{1}{2}\tau^2} \right| d\tau \\ & \leq \int_{|\tau| \leq \omega\sqrt{n}} \exp \left(-\frac{\tau^2}{4} \right) \left(\frac{\rho|\tau|^3}{\sqrt{n}} + \frac{\mu_3^2\tau^6}{72ns^6} \right) d\tau \\ & = \frac{\rho}{\sqrt{n}} \int_{|\tau| \leq \omega\sqrt{n}} \exp \left(-\frac{\tau^2}{4} \right) |\tau|^3 d\tau + \frac{\mu_3^2}{72ns^6} \int_{|\tau| \leq \omega\sqrt{n}} \exp \left(-\frac{\tau^2}{4} \right) \tau^6 d\tau, \end{aligned}$$

where both the first integral and the second integral are finite, and ρ is arbitrarily small; use Corollary 5, to obtain

$$\int_{|\tau| \leq \omega\sqrt{n}} \left| (\varphi^{a_n}(\tau/\sqrt{n}))^n - e^{-\frac{1}{2}\tau^2} - \frac{\mu_3}{6\sqrt{n}s^3}(i\tau)^3 e^{-\frac{1}{2}\tau^2} \right| d\tau = o\left(\frac{1}{\sqrt{n}}\right).$$

This gives (34), and therefore we obtain

$$\left| \bar{\pi}_n^{a_n}(x) - \phi(x) - \frac{\mu_3}{6\sqrt{n}s^3}(x^3 - 3x)\phi(x) \right| = o\left(\frac{1}{\sqrt{n}}\right),$$

which concludes the proof.

6.2 Proof of Theorem 8

It is well known and easily checked that the conditional density $p(X_1^k = y_1^k | S_1^n = na_n)$ is invariant under any i.i.d sampling scheme in the family of densities π^α as α belongs to $Im(X_1)$ (commonly called tilting change of measure). Namely

$$p(X_1^k = y_1^k | S_1^n = na_n) = \pi^\alpha(X_1^k = y_1^k | S_1^n = na_n)$$

where on the LHS the X_i 's are sampled i.i.d. under p and on the RHS they are sampled i.i.d. under π^α .

Using Bayes formula, it thus holds

$$\begin{aligned} p(X_1 = y_1 | S_1^n = na_n) &= \pi^m(X_1 = y_1 | S_1^n = na_n) \\ &= \pi^m(X_1 = y_1) \frac{\pi^m(S_2^n = na_n - y_1)}{\pi^m(S_1^n = na_n)} \\ &= \frac{\sqrt{n}}{\sqrt{n-1}} \pi^m(X_1 = y_1) \frac{\widetilde{\pi}_{n-1}\left(\frac{m-y_1}{s\sqrt{n-1}}\right)}{\widetilde{\pi}_n(0)}, \end{aligned} \quad (38)$$

where $\widetilde{\pi}_{n-1}$ is the normalized density of S_2^n under i.i.d. sampling with the density π^{a_n} ; correspondingly, $\widetilde{\pi}_n$ is the normalized density of S_1^n under the same sampling. Note that a r.v. with density π^{a_n} has expectation m and variance s^2 . Perform a third-order Edgeworth expansion of $\widetilde{\pi}_{n-1}(z)$, using Theorem 6. It follows

$$\widetilde{\pi}_{n-1}(z) = \phi(z) \left(1 + \frac{\mu_3}{6s^3\sqrt{n-1}}(z^3 - 3z) \right) + o\left(\frac{1}{\sqrt{n}}\right),$$

The approximation of $\widetilde{\pi}_n(0)$ is

$$\widetilde{\pi}_n(0) = \phi(0) \left(1 + o\left(\frac{1}{\sqrt{n}}\right) \right).$$

Hence (38) becomes

$$\begin{aligned} p(X_1 = y_1 | S_1^n = na_n) &= \frac{\sqrt{n}}{\sqrt{n-1}} \pi^m(X_1 = y_1) \frac{\phi(z)}{\phi(0)} \left[1 + \frac{\mu_3}{6s^3\sqrt{n-1}}(z^3 - 3z) + o\left(\frac{1}{\sqrt{n}}\right) \right] \\ &= \frac{\sqrt{2\pi n}}{\sqrt{n-1}} \pi^m(X = y_1) \phi(z) (1 + R_n + o(1/\sqrt{n})), \end{aligned} \quad (39)$$

where

$$R_n = \frac{\mu_3}{6s^3\sqrt{n-1}}(z^3 - 3z).$$

Under condition (16), by Corollary (5), $\mu_3/s^3 \rightarrow 0$. This yields

$$R_n = o(1/\sqrt{n}),$$

which gives

$$p(X_1 = y_1 | S_1^n = na_n) = \pi^m(X = y_1)(1 + o(1/\sqrt{n}))$$

as claimed.

6.3 Proof of Proposition 11

Denote

$$z_i := \frac{m_i - y_{i+1}}{s_i \sqrt{n - i - 1}}$$

where

$$s_i^2 := s^2(t_i).$$

We first state a Lemma pertaining to the order of magnitude of z_i . The proof of this Lemma is in the next Subsection

Lemma 31 *Assume that $h(x) \in \mathcal{R}$. Let t_i be defined by (19). Assume that $a_n \rightarrow \infty$ as $n \rightarrow \infty$ and that (16) holds. Then as $n \rightarrow \infty$*

$$\lim_{n \rightarrow \infty} \sup_{0 \leq i \leq k-1} z_i = 0, \quad \text{and} \quad \sup_{0 \leq i \leq k-1} z_i^2 = o\left(\frac{1}{\sqrt{n}}\right).$$

We turn to the proof of Proposition 11.

It holds by Bayes formula,

$$p_{a_n}(y_1^k) = \prod_{i=0}^{k-1} p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - s_i^i).$$

Using the invariance of the conditional distributions under the tilting it holds, for any i between 1 and $k - 1$

$$\begin{aligned} p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - S_i^i) &= \frac{\sqrt{2\pi(n-i)}}{\sqrt{n-i-1}} \pi^{m_i}(X_{i+1} = y_{i+1}) \phi(z_i) (1 + o(1/\sqrt{n})) \\ &= \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}(X_{i+1} = y_{i+1}) (1 - z_i^2/2 + o(z_i^2)) (1 + o(1/\sqrt{n})), \end{aligned}$$

where we used a Taylor expansion in the second equality. Using once more Lemma 31, under conditions (16), we have as $a_n \rightarrow \infty$

$$z_i^2 = o(1/\sqrt{n}).$$

Hence we get

$$p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - s_1^i) = \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}(X_{i+1} = y_{i+1}) (1 + o(1/\sqrt{n})),$$

which yields

$$\begin{aligned} p(X_1^k = y_1^k | S_1^n = na_n) &= \prod_{i=0}^{k-1} \left(\frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{m_i}(X_{i+1} = y_{i+1}) (1 + o(1/\sqrt{n})) \right) \\ &= \prod_{i=0}^{k-1} \pi^{m_i}(X_{i+1} = y_{i+1}) \prod_{i=0}^{k-1} \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \prod_{i=0}^{k-1} \left(1 + o\left(\frac{1}{\sqrt{n}}\right) \right) \\ &= \left(1 + o\left(\frac{1}{\sqrt{n}}\right) \right) \prod_{i=0}^{k-1} \pi^{m_i}(X_{i+1} = y_{i+1}), \end{aligned}$$

The proof is completed.

6.4 Proof of Lemma 31

When $n \rightarrow \infty$, it holds

$$z_i \sim m_i / (s_i \sqrt{n}).$$

From Theorem 4, it holds

$$z_i \sim \frac{\psi(t_i)}{\sqrt{n\psi'(t_i)}}.$$

Since $m_i \sim m_k$ as $n \rightarrow \infty$, it holds

$$m_i \sim \psi(t_k).$$

Hence

$$\psi(t_i) \sim \psi(t_k).$$

Case 1: if $h(x) \in R_\beta$. Hence

$$h'(x) = x^{\beta-1} l_0(x) (\beta + \epsilon(x)).$$

Set $x = \psi(t)$; we get

$$h'(\psi(t)) = \psi(t)^{\beta-1} l_0(\psi(t)) (\beta + \epsilon(\psi(t))).$$

Notice that $\psi'(t) = 1/h'(\psi(t))$; we obtain

$$\frac{\psi'(t_i)}{\psi'(t_k)} = \frac{h'(\psi(t_k))}{h'(\psi(t_i))} = \frac{(\psi(t_k))^{\beta-1} l_0(\psi(t_k)) (\beta + \epsilon(\psi(t_k)))}{(\psi(t_i))^{\beta-1} l_0(\psi(t_i)) (\beta + \epsilon(\psi(t_i)))} \rightarrow 1,$$

where we use the slowly varying propriety of l_0 . Thus it holds

$$\psi'(t_i) \sim \psi'(t_k),$$

which yields

$$z_i \sim \frac{\psi(t_k)}{\sqrt{n\psi'(t_k)}}.$$

Hence we have under condition (16)

$$z_i^2 \sim \frac{\psi(t_k)^2}{n\psi'(t_k)} = \frac{\psi(t_k)^2}{\sqrt{n}\psi'(t_k)} \frac{1}{\sqrt{n}} = o\left(\frac{1}{\sqrt{n}}\right),$$

which implies further that $z_i \rightarrow 0$.

Case 2: if $h(x) \in R_\infty$. It holds $m(t_k) \geq m(t_i)$ as $n \rightarrow \infty$. Since the function $t \rightarrow m(t)$ is increasing, we have

$$t_i \leq t_k.$$

The function $t \rightarrow \psi'(t)$ is decreasing, since

$$\psi''(t) = -\frac{\psi(t)}{t^2}\epsilon(t)(1+o(1)) < 0 \quad \text{as } t \rightarrow \infty.$$

Therefore as $n \rightarrow \infty$

$$\psi'(t_i) \geq \psi'(t_k) > 0,$$

which yields

$$z_i \sim \frac{\psi(t_i)}{\sqrt{n\psi'(t_i)}} \leq \frac{2\psi(t_k)}{\sqrt{n\psi'(t_k)}},$$

hence we have

$$z_i^2 \leq \frac{4\psi(t_k)^2}{n\psi'(t_k)} = \frac{4\psi(t_k)^2}{\sqrt{n}\psi'(t_k)} \frac{1}{\sqrt{n}} = o\left(\frac{1}{\sqrt{n}}\right),$$

where the last step holds from condition (16). Further it holds $z_i \rightarrow 0$.

This closes the proof of the Lemma.

6.5 Proof of Lemma 22

Case 1: if $h(t) \in R_\beta$. By Theorem 4, it holds $s^2 \sim \psi'(t)$ with $\psi(t) \sim t^{1/\beta}l_1(t)$, where l is some slowly varying function. Consider $\psi'(t) = 1/h'(\psi(t))$, hence

$$\begin{aligned} \frac{1}{s^2} &\sim h'(\psi(t)) = \psi(t)^{\beta-1}l_0(\psi(t))(\beta + \epsilon(\psi(t))) \\ &\sim \beta t^{1-1/\beta}l_1(t)^{\beta-1}l_0(\psi(t)) = o(t), \end{aligned}$$

where $l_0 \in R_0$. This implies for any $u \in K$

$$\frac{u}{s} = o(\sqrt{t}),$$

which yields, using (9)

$$\begin{aligned} \frac{s^2(t+u/s)}{s^2} &\sim \frac{\psi'(t+u/s)}{\psi'(t)} = \frac{\psi(t)^{\beta-1}l_0(\psi(t))(\beta + \epsilon(\psi(t)))}{(\psi(t+u/s))^{\beta-1}l_0(\psi(t+u/s))(\beta + \epsilon(\psi(t+u/s)))} \\ &\sim \frac{\psi(t)^{\beta-1}}{\psi(t+u/s)^{\beta-1}} \sim \frac{t^{1-1/\beta}l_1(t)^{\beta-1}}{(t+u/s)^{1-1/\beta}l_1(t+u/s)^{\beta-1}} \rightarrow 1. \end{aligned}$$

Case 2: if $h(t) \in R_\infty$. Then $\psi(t) \in \widetilde{R}_0$, hence it holds

$$\frac{1}{st} \sim \frac{1}{t\sqrt{\psi'(t)}} = \sqrt{\frac{1}{t\psi(t)\epsilon(t)}} \rightarrow 0,$$

which last step holds from condition (12). Hence for any $u \in K$, we get as $n \rightarrow \infty$

$$\frac{u}{s} = o(t),$$

thus using the slowly varying propriety of $\psi(t)$ we have

$$\begin{aligned} \frac{s^2(t+u/s)}{s^2} &\sim \frac{\psi'(t+u/s)}{\psi'(t)} = \frac{\psi(t+u/s)\epsilon(t+u/s)}{t+u/s} \frac{t}{\psi(t)\epsilon(t)} \\ &\sim \frac{\epsilon(t+u/s)}{\epsilon(t)} = \frac{\epsilon(t) + O(\epsilon'(t)u/s)}{\epsilon(t)} \rightarrow 1, \end{aligned} \quad (40)$$

where we used a Taylor expansion in the second line, and where the last step holds from condition (12). This completes the proof.

6.6 Proof of Theorem 12

Making use of

$$p(X_1^k = y_1^k | S_1^n = na_n) = \prod_{i=0}^{k-1} p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - s_1^i),$$

and using the tilted density π^{a_n} instead of π^{m_i} it holds

$$p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - s_1^i) = \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{a_n}(X_{i+1} = y_{i+1}) \frac{\widetilde{\pi_{n-i-1}\left(\frac{(i+1)a_n - s_1^{i+1}}{s\sqrt{n-i-1}}\right)}}{\widetilde{\pi_{n-i}\left(\frac{ia_n - s_1^i}{s\sqrt{n-i}}\right)}}, \quad (41)$$

where $\widetilde{\pi_{n-i-1}}$ is the normalized density of S_{i+2}^n under i.i.d. sampling with π^{a_n} . Correspondingly, denote $\widetilde{\pi_{n-i}}$ the normalized density of S_{i+1}^n under the same sampling. Write

$$z_i = \frac{ia_n - s_1^{i-1}}{s\sqrt{n-i+1}}.$$

By Theorem 6 a third-order Edgeworth expansion yields

$$\widetilde{\pi_{n-i-1}}(z_i) = \phi(z_i) (1 + R_n^i) + o\left(\frac{1}{\sqrt{n}}\right),$$

where

$$R_n^i = \frac{\mu_3}{6s^3\sqrt{n-i-1}}(z_i^3 - 3z_i).$$

Accordingly

$$\widetilde{\pi_{n-i}}(z_{i-1}) = \phi(z_{i-1}) (1 + R_n^{i-1}) + o\left(\frac{1}{\sqrt{n}}\right).$$

When $a_n \rightarrow \infty$, using Theorem 4, it holds

$$\begin{aligned} \sup_{0 \leq i \leq k-1} z_i^2 &\sim \frac{(i+1)^2 a_n^2}{s^2 n} \leq \frac{2k^2 a_n^2}{s^2 n} = \frac{2k^2 (m(t))^2}{s^2 n} \\ &\sim \frac{2k^2 (\psi(t))^2}{\psi'(t)n} = \frac{2k^2 (\psi(t))^2}{\sqrt{n}\psi'(t)} \frac{1}{\sqrt{n}} = o\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \quad (42)$$

where the last step holds under condition (16). Hence it holds $z_i \rightarrow 0$ for $0 \leq i \leq k-1$ as $a_n \rightarrow \infty$, and by Corollary 5, $\mu_3/s^3 \rightarrow 0$; Hence

$$R_n^i = o(1/\sqrt{n}) \text{ and } R_n^{i-1} = o(1/\sqrt{n}).$$

We thus get

$$\begin{aligned} p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - s_1^i) &= \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{a_n}(X_{i+1} = y_{i+1}) \frac{\phi(z_i)}{\phi(z_{i-1})} (1 + o(1/\sqrt{n})) \\ &= \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{a_n}(X_{i+1} = y_{i+1}) (1 - (z_i^2 - z_{i-1}^2)/2 + o(z_i^2 - z_{i-1}^2)) (1 + o(1/\sqrt{n})), \end{aligned}$$

where we used a Taylor expansion in the second equality. Using (42), we have as $a_n \rightarrow \infty$

$$|z_i^2 - z_{i-1}^2| = o(1/\sqrt{n}),$$

from which

$$p(X_{i+1} = y_{i+1} | S_{i+1}^n = na_n - s_1^i) = \frac{\sqrt{n-i}}{\sqrt{n-i-1}} \pi^{a_n}(X_{i+1} = y_{i+1}) (1 + o(1/\sqrt{n})),$$

which yields

$$\begin{aligned} p(X_1^k = y_1^k | S_1^n = na_n) &= \prod_{i=0}^{k-1} \left(\pi^{a_n}(X_{i+1} = y_{i+1}) \sqrt{\frac{n}{n-k}} \right) \prod_{i=0}^{k-1} \left(1 + o\left(\frac{1}{\sqrt{n}}\right) \right) \\ &= \left(1 + o\left(\frac{1}{\sqrt{n}}\right) \right) \prod_{i=0}^{k-1} \pi^{a_n}(X_{i+1} = y_{i+1}). \end{aligned}$$

This completes the proof.

6.7 Proof of Lemma 27

For a density $p(x)$ defined in as in (5), we show that $g(x)$ is a convex function when x is large. If $h(x) \in R_\beta$, for x large

$$g''(x) = h'(x) = \frac{h(x)}{x} (\beta + \epsilon(x)) > 0.$$

If $h(x) \in R_\infty$, its reciprocal function $\psi(x) \in \widetilde{R}_0$. Set $x := \psi(v)$. Then

$$g''(x) = h'(x) = \frac{1}{\psi'(v)} = \frac{v}{\psi(v)\epsilon(v)} > 0,$$

where the inequality holds since $\epsilon(v) > 0$ when v is large enough. Hence $g(x)$ is convex for large x . Therefore, the density $p(x)$ with $h(x) \in \mathcal{R}$ satisfies the conditions of Theorem 6.2.1 in [14]. Denote by p_n the density of $\bar{X} = (X_1 + \dots + X_n)/n$. We obtain from formula (2.2.6) of [14], using a third order Edgeworth expansion

$$P(S_1^n \geq na_n) = \frac{\Phi(t)^n \exp(-nta_n)}{\sqrt{nts(t)}} (B_0(\lambda_n)) + O\left(\frac{\mu_3(t)}{6\sqrt{n}s^3(t)} B_3(\lambda_n)\right),$$

where $\lambda_n = \sqrt{nts(t)}$, $B_0(\lambda_n)$ and $B_3(\lambda_n)$ are defined by

$$B_0(\lambda_n) = \frac{1}{\sqrt{2\pi}} \left(1 - \frac{1}{\lambda_n^2} + o\left(\frac{1}{\lambda_n^2}\right) \right), \quad B_3(\lambda_n) \sim -\frac{3}{\sqrt{2\pi}\lambda_n}.$$

We show that as $a_n \rightarrow \infty$

$$\frac{1}{\lambda_n^2} = o\left(\frac{1}{n}\right). \quad (43)$$

Since $n/\lambda_n^2 = 1/(t^2s^2(t))$, (43) is equivalent to show that

$$t^2s^2(t) \longrightarrow \infty.$$

By Theorem 4, $m(t) \sim \psi(t)$ and $s^2(t) \sim \psi'(t)$; combined with $m(t) = a_n$, it holds $t \sim h(a_n)l_1(a_n)$, where l_1 is some slowly varying function. If $h \in R_\beta$, notice that

$$\psi'(t) = \frac{1}{h'(\psi(t))} = \frac{\psi(t)}{h(\psi(t))(\beta + \epsilon(\psi(t)))} \sim \frac{a_n}{h(a_n)(\beta + \epsilon(\psi(t)))};$$

hence

$$t^2 s^2(t) \sim h(a_n)^2 l_1(a_n)^2 \frac{a_n}{h(a_n)(\beta + \epsilon(\psi(t)))} = \frac{a_n h(a_n) l_1(a_n)^2}{\beta + \epsilon(\psi(t_n))} \rightarrow \infty.$$

If $h \in R_\infty$, then $\psi(t) \in \widetilde{R}_0$, thus

$$t^2 s^2(t) \sim t^2 \frac{\psi(t)\epsilon(t)}{t} = t\psi(t)\epsilon(t) \rightarrow \infty,$$

Summing up we have proved that

$$B_0(\lambda_n) = \frac{1}{\sqrt{2\pi}} \left(1 + o\left(\frac{1}{n}\right) \right).$$

By (43), λ_n goes to ∞ as $a_n \rightarrow \infty$; this implies further that $B_3(\lambda_n) \rightarrow 0$. On the other hand, by Corollary 5 it holds $\mu_3/s^3 \rightarrow 0$. Hence we obtain

$$P(S_1^n \geq na_n) = \frac{\Phi(t)^n \exp(-nta_n)}{\sqrt{2\pi nts(t)}} \left(1 + o\left(\frac{1}{\sqrt{n}}\right) \right),$$

which gives (24). By Jensen's Theorem 6.2.1 ([14]) and formula (2.2.4) in [14] it follows uniformly in τ

$$p(S_1^n/n = \tau) = \frac{\sqrt{n}\Phi(t_\tau)^n \exp(-nt_\tau\tau)}{\sqrt{2\pi s(t_\tau)}} \left(1 + o\left(\frac{1}{\sqrt{n}}\right) \right),$$

which, together with $p(S_1^n = n\tau) = (1/n)p(S_1^n/n = \tau)$, gives (25).

6.8 Proof of Theorem 28

It holds

$$\begin{aligned} p_{A_n}(y_1) &= \int_{a_n}^{\infty} p(X_1 = y_1 | S_1^n = n\tau) p(S_1^n = n\tau | S_1^n \geq na_n) d\tau \\ &= \frac{p(X_1 = y_1)}{P(S_1^n \geq na_n)} \int_{a_n}^{\infty} p(S_2^n = n\tau - y_1) d\tau \\ &= \left(1 + \frac{P_2}{P_1} \right) \frac{p(X_1 = y_1)}{P(S_1^n \geq na_n)} \int_{a_n}^{a_n + \eta_n} p(S_2^n = n\tau - y_1) d\tau \\ &= \left(1 + \frac{P_2}{P_1} \right) \int_{a_n}^{a_n + \eta_n} p(X_1 = y_1 | S_1^n = n\tau) p(S_1^n = n\tau | S_1^n \geq na_n) d\tau \end{aligned} \tag{44}$$

where the second equality is obtained by Bayes formula, and

$$P_1 = \int_{a_n}^{a_n + \eta_n} p(S_2^n = n\tau - y_1) d\tau,$$

$$P_2 = \int_{a_n + \eta_n}^{\infty} p(S_2^n = n\tau - y_1) d\tau.$$

We show that P_2 is infinitely small with respect to P_1 . Indeed

$$P_2 = \frac{1}{n} P(S_2^n \geq n(a_n + \eta_n) - y_1) = \frac{1}{n} P(S_2^n \geq (n-1)c_n),$$

$$P_1 + P_2 = \frac{1}{n} P(S_2^n \geq na_n - y_1) = \frac{1}{n} P(S_2^n \geq (n-1)d_n),$$

where $c_n = (n(a_n + \eta_n) - y_1)/(n-1)$ and $d_n = (na_n - y_1)/(n-1)$. Denote $t_{c_n} = m^{-1}(c_n)$ and $t_{d_n} = m^{-1}(d_n)$. Using Lemma 27, it holds

$$\frac{P_2}{P_1 + P_2} = \left(+o\left(\frac{1}{\sqrt{n}}\right) \right) \frac{t_{d_n} s(t_{d_n})}{t_{c_n} s(t_{c_n})} \exp(-(n-1)(I(c_n) - I(d_n))).$$

Using the convexity of the function I , it holds

$$\begin{aligned} \exp(-(n-1)I(c_n) - I(d_n)) &\leq \exp-(n-1)(c_n - d_n)m^{-1}(d_n) \\ &= \exp-n\eta_n m^{-1}(d_n). \end{aligned}$$

The function $u \rightarrow m^{-1}(u)$ is increasing. Since $d_n \geq a_n$ as $a_n \rightarrow \infty$, it holds $m^{-1}(d_n) \geq m^{-1}(a_n)$; hence $\exp-(n-1)(I(c_n) - I(d_n)) \leq \exp-n\eta_n m^{-1}(a_n) \rightarrow 0$. We now show that

$$\frac{t_{d_n} s(t_{d_n})}{t_{c_n} s(t_{c_n})} \rightarrow 1.$$

By definition, $c_n/d_n \rightarrow 1$ as $a_n \rightarrow \infty$. If $h \in R_\beta$, it holds

$$\left(\frac{t_{d_n} s(t_{d_n})}{t_{c_n} s(t_{c_n})} \right)^2 \sim \left(\frac{d_n h(d_n)}{\beta + \epsilon(\psi(d_n))} \right)^2 \left(\frac{\beta + \epsilon(\psi(c_n))}{c_n h(c_n)} \right)^2 \sim \left(\frac{h(d_n)}{h(c_n)} \right)^2 \rightarrow 1.$$

If $h \in R_\infty$,

$$t^2 s^2(t) \sim t\psi(t)\epsilon(t),$$

hence

$$\left(\frac{t_{d_n} s(t_{d_n})}{t_{c_n} s(t_{c_n})} \right)^2 \sim \frac{d_n \psi(d_n) \epsilon(d_n)}{c_n \psi(c_n) \epsilon(c_n)} \sim \frac{\epsilon(d_n)}{\epsilon(c_n)} = \frac{\epsilon(c_n - n\eta_n/(n-1))}{\epsilon(c_n)} \rightarrow 1,$$

where last step holds by using the same argument as in the second line of (40). We obtain

$$\frac{P_2}{P_1} = o(1).$$

Therefore $p_{A_n}(y_1)$ can be approximated by

$$p_{A_n}(y_1) = (1 + o(1)) \int_{a_n}^{a_n + \eta_n} p(X_1 = y_1 | S_1^n = n\tau) p(S_1^n = n\tau | S_1^n \geq na_n) d\tau.$$

By Lemma 27, it follows that uniformly when $\tau \in [a_n, a_n + \eta_n]$

$$\begin{aligned} p(S_1^n = n\tau | S_1^n \geq na_n) &= \frac{p(S_1^n = n\tau)}{P(S_1^n \geq na_n)} \\ &= \left(1 + o\left(\frac{1}{\sqrt{n}}\right)\right) \frac{ts(t)}{s(t_\tau)} \exp(-n(I(\tau) - I(a_n))), \end{aligned} \quad (45)$$

We now turn back to (44) and note that under the appropriate condition (16) or (17) the corresponding approximating density π^τ or g_τ can be seen to hold uniformly on τ in $(a_n, a_n + \eta_n)$. Inserting (45) into (44), we complete the proof of Theorem 28 inserting the corresponding local result.

References

- [1] Balkema, A. A.; Klüppelberg, C.; Resnick, S. I. Densities with Gaussian tails. Proc. London Math. Soc. (3) 66 (1993), no. 3, 568–588.
- [2] Bar-Lev, S K.; Bshouty, D; Enis, P On polynomial variance functions. Probab. Theory Related Fields 94 (1992), no. 1, 69–82.
- [3] Barndorff-Nielsen, O., “Information and exponential families in statistical theory,” Wiley Series in Probability and Mathematical Statistics. Chichester: John Wiley & Sons, (1978).
- [4] Bingham, N.H., Goldie, C.M., Teugels, J.L., “Regular Variation,” Cambridge University Press, Cambridge, (1987).
- [5] Bhattacharya, R. N., Rao, Ranga R., “Normal approximation and asymptotic expansions,” Society for Industrial and Applied Mathematics, Philadelphia, (2010).
- [6] Biret ,M., Broniatowski, M.,Cao, Z, A sharp Abelian Theorem for the Laplace transform, Limit Theorems in Mathematical Statistics, Ed: M;Hallin, D.Mason, D.Pfeiffer, Springer 2015.
- [7] Broniatowski, M., Cao, Z. , Light tails: all summands are large when the empirical mean is large. Extremes 17 (2014), no. 2, 305–336.

- [8] Broniatowski, M., Caron, V. Long runs under a conditional limit distribution. *Ann. Appl. Probab.* 24 (2014), no. 6, 2246–2296.
- [9] Csiszár, I. Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Probab.* 12 (1984), no. 3, 768–793.
- [10] Dembo, A. and Zeitouni, O., Refinements of the Gibbs conditioning principle, *Probab. Theory Related Fields* 104 1–14, (1996).
- [11] Diaconis, P., Freedman, D., Conditional Limit Theorems for Exponential Families and Finite Versions of de Finetti’s Theorem, *Journal of Theoretical Probability*, Vol. 1, No. 4, (1988).
- [12] Feigin, P. D., Yashchin, E. On a strong Tauberian result. *Z. Wahrsch. Verw. Gebiete* 65 (1983), no. 1, 35–48.
- [13] Feller, W. “An introduction to probability theory and its applications,” Vol. 2, second edition, John Wiley and Sons Inc., New York, (1971).
- [14] Jensen, J. L. “Saddlepoint approximations,” Oxford Statistical Science Series, 16. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, (1995).
- [15] Jørgensen, B; Martínez, J R. Tauber theory for infinitely divisible variance functions. *Bernoulli* 3 (1997), no. 2, 213–224.
- [16] Juszcak, D., Nagaev, A. V., Local large deviation theorem for sums of i.i.d. random vectors when the Cramer condition holds in the whole space, *Probability and Mathematical Statistics*, Vol. 24, (2004), pp. 297-320.