Probability Surveys
Vol. 9 (2012) 439–479
ISSN: 1549-5787
DOI: 10.1214/12-PS198

Quantile coupling inequalities and their applications

David M. Mason*

Statistics Program
University of Delaware
Newark, DE 19716
e-mail: davidm@udel.edu
url: http://ag.udel.edu/frec/faculty/mason.htm

and

Harrison H. Zhou†

Department of Statistics
Yale University
New Haven, CT 06511
e-mail: huibin.zhou@yale.edu
url: http://www.stat.yale.edu/~hz68

Abstract: This is partly an expository paper. We prove and highlight a quantile inequality that is implicit in the fundamental paper by Komlós, Major, and Tusnády [31] on Brownian motion strong approximations to partial sums of independent and identically distributed random variables. We also derive a number of refinements of this inequality, which hold when more assumptions are added. A number of examples are detailed that will likely be of separate interest. We especially call attention to applications to the asymptotic equivalence theory of nonparametric statistical models and nonparametric function estimation.

AMS 2000 subject classifications: Primary 62E17; secondary 62B15, 62G05.
Keywords and phrases: Quantile coupling, large deviation, KMT construction, asymptotic equivalence, function estimation.

Received March 2012.

1. Introduction

Komlós, Major, and Tusnády [KMT] [31, 32] approximations to the partial sum and empirical processes are two of the most important results in probability over the last forty years. In particular, they proved the following powerful Gaussian coupling to partial sums [PS] of i.i.d. random variables. (We shall use the words *approximation* and *coupling* almost interchangeably.)

*The research of David Mason was supported in part by NSF Grant DMS-05-03908 and the Unit for Busimess Mathematics and Informatics at North-West University, South Africa, where he holds the position as Extraordinary Professor.

†The research of Harrison Zhou was supported in part by NSF Career Award DMS-0645676 and NSF FRG Grant DMS-0854975.

Theorem [PS]. *Let X be a random variable with mean 0 and variance $0 < \sigma^2 < \infty$. Also assume that $E\exp(a|X|) < \infty$ for some $a > 0$. Then on the same probability space there exist i.i.d. X random variables $X_1, X_2, \ldots,$ and i.i.d. standard normal random variables $Z_1, Z_2, \ldots,$ such that for positive constants C, D and λ for all $x \in \mathbb{R}$ and $n \geq 1$,*

$$P\left\{\max_{1\le k\le n}\left|\sigma^{-1}\sum_{i=1}^{k}X_i-\sum_{i=1}^{k}Z_i\right|>D\log n+x\right\}\le C\exp\left(-\lambda x\right). \tag{1}$$

This is Theorem 1 of KMT [32]. The original version given in Theorem 1 of KMT [31] is stated under added conditions.

One of the key tools needed in its proof was a quantile inequality. To describe it let us introduce some notation. Let $\{Y_n\}_{n\geq 1}$ be a sequence of random variables and for each integer $n \geq 1$ let

$$F_n(x) = P\{Y_n \le x\}, \text{ for } x \in \mathbb{R}, \tag{2}$$

denote the cumulative distribution function [cdf] of Y_n. Its *inverse distribution function* or *quantile function* is defined by

$$H_n(s) = \inf\{x : F_n(x) \ge s\}, \text{ for } s \in (0,1). \tag{3}$$

Let Z denote a standard normal random variable, Φ be its cdf and ϕ its density function. Since $\Phi(Z) =_d U$, we see that for each integer $n \geq 1$,

$$H_n(\Phi(Z)) =_d Y_n.$$

For this reason, we shall from now on write for convenience

$$H_n(\Phi(Z)) = Y_n. \tag{4}$$

Consider now the special case of $\{Y_n\}_{n\geq 1}$ such that for each $n \geq 1$,

$$Y_n =_d \sum_{i=1}^{n} X_i/\left(\sigma\sqrt{n}\right), \tag{5}$$

where $X_1, X_2, \ldots,$ are i.i.d. X satisfying the conditions of Theorem [PS]. Fundamental to the proof of Theorem [PS] is the following quantile inequality, which is implicit in the proof of Theorem 1 of KMT [31].

Proposition [KMT]. *Assume that $X_1, X_2, \ldots,$ are i.i.d. X satisfying the conditions of Theorem [PS]. Then there exist a $0 < D < \infty$ and an $0 < \eta < \infty$ such that for all integers $n \geq 1$, whenever Y_n is as in (5) and (4), and*

$$|Y_n| \le \eta\sqrt{n},$$

we have

$$|Y_n - Z| \le \frac{DY_n^2}{\sqrt{n}} + \frac{D}{\sqrt{n}}.$$

For a multidimensional version of Proposition [KMT] consult Einmahl [19].

We shall soon show that if additional assumptions are imposed on X that this inequality can be improved, in particular when X is symmetric and its distribution has a nonzero absolutely continuous component. Refer to Theorem 3 and Proposition 5 below for details.

We shall also see that this inequality leads to a coupling of Y_n and Z such that for suitable constants $C > 0$ and $\lambda > 0$

$$P\left\{\sqrt{n}\,|Y_n - Z| > z\right\} \leq C\exp\left(-\lambda z\right), \text{ for all } z \geq 0, \tag{6}$$

which via Lemma A1 of Berkes and Philipp [4] implies that for each integer $n \geq 1$ there exist $X_1, \ldots, X_n$ i.i.d. X and i.i.d. standard normal random variables $Z_1, \ldots, Z_n$ such that on a suitable probability space for all $z \geq 0$

$$P\left\{\sqrt{n}\left|\sigma^{-1}\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Z_i\right| > z\right\} \leq C\exp\left(-\lambda z\right).$$

KMT [31] also stated the following Brownian bridge coupling to the uniform empirical process α_n, along with an outline of its proof. But first, here is the definition of α_n. Let U_1, $U_2, \ldots,$ be a sequence of independent Uniform $(0,1)$ random variables. For each integer $n \geq 1$ let

$$G_n(t) = n^{-1}\sum_{i=1}^{n} 1\{U_i \leq t\},\ t \in \mathbb{R},$$

denote the *empirical distribution function* based on $U_1, \ldots, U_n$. The *uniform empirical process [EP]* α_n is the process

$$\alpha_n(t) = \sqrt{n}\{G_n(t) - t\},\ t \in [0,1]\,. \tag{7}$$

Theorem [EP]. *There exists a probability space* $(\Omega, \mathcal{A}, P)$ *with a sequence of independent Uniform*$[0,1]$ *random variables* $U_1, U_2, \ldots,$ *a sequence of Brownian bridges* $B_1, B_2, \ldots,$ *and positive constants* a, b *and* c *such that for all* $n \geq 1$ *and* $x \in \mathbb{R}$,

$$P\left\{\sup_{0\leq t\leq 1}|\sqrt{n}\left\{\alpha_n(t) - B_n(t)\right\}| \geq a\log n + x\right\} \leq b\exp(-cx).$$

Mason and van Zwet [38], Major [36] and Mason [37] have published the details of the proof of Theorem [EP] based on Proposition [KMT] as it applies to

$$Y_n =_d \frac{2S_n - n}{\sqrt{n}}, \tag{8}$$

where S_n is a Binomial random variable with parameters n and $1/2$. A proof of Theorem [EP] can also be obtained using a binomial inequality due to Tusnády (Proposition [T]) [55]. This inequality is often referred to as the Tusnády lemma, which for comparison we state here.

Proposition [T]. *For all integers $n \geq 1$, with Y_n as in (8) and in (4)*

$$|Y_n| \leq \frac{2}{\sqrt{n}} + |Z|$$

and

$$|Y_n - Z| \leq \frac{2}{\sqrt{n}} + \frac{Z^2}{4\sqrt{n}}.$$

Tusnády [55] did not provide a fully detailed proof of his lemma. In fact, M. Csörgő and Révész [12] remarked in their monograph on strong approximations, "Although the proof of the inequality is elementary, it is not simple. It will not be given here however." When Bretagnolle and Massart [6] published a complete proof of the Tusnády lemma, it indeed was not simple. Other proofs of the Tusnády lemma can be found in M. Csörgő and Horváth [11], Dudley [17], Massart [40] and Lawler and Trujillo Ferreras [33]. Carter and Pollard [9] improved upon the Tusnády inequality. More specifically, they showed that with Y_n as in (8) and (4)

$$|Y_n - Z| \leq \frac{C}{\sqrt{n}} + \frac{C\,|Z|^3}{n}, \text{ whenever } |Y_n| \leq \varepsilon\sqrt{n} \tag{9}$$

for some $C, \varepsilon > 0$. Bretagnolle and Massart [6], Csörgő and Horváth [11] and Dudley [17]) also give proofs of Theorem [EP] based on the Tusnády lemma. It is sometimes thought that the Tusnády lemma is indispensable to its proof. However this is not the case. As pointed out above, its original proof as sketched in KMT [31] was based on the binomial special case of Proposition [KMT]. Clearly the quantile coupling of a standardized sum of i.i.d. Bernoulli(1/2) with a normal random variable lies at the heart of KMT construction for the empirical process.

In the last decade the KMT construction has played a key role in the progress of the asymptotic equivalence of experiment theory. Nussbaum [41] made a remarkable breakthrough in asymptotic equivalence theory using KMT. He established the asymptotic equivalence of density estimation and Gaussian white noise under a Hölder smoothness condition. A major step toward the proof of this equivalence result is the functional KMT construction for the empirical process by Koltchinskii [30]. His construction relies on the Tusnády lemma. The main consequence of this result is that an asymptotically optimal result in one of these nonparametric models automatically yields an analogous result in the other model.

Our paper is largely expository. Its two goals are to spotlight and prove a basic quantile inequality that is implicit in KMT [31], as well as establish the improvements that we alluded to above, and then show how they can be used to obtain a number of interesting couplings of a sequence of random variables Y_n to a standard normal random variable Z and describe their applications in probability and statistics. As a by-product, we get Proposition [KMT] and (9) as special cases of formally more general results. In an applications section we shall also describe how refinements to the KMT quantile inequality lead to advances in asymptotic equivalence of experiment theory. It is hoped that this paper helps to make these quantile inequalities known to a wider mathematical and statistical audience.

Our paper is organized as follows. In Section 2, we state our basic quantile couplings, then in Section 3 we discuss examples of their use in probability theory and the theory of statistical experiments. Section 4 is devoted to proofs. Appendices A, B and C provide additional information for the interested reader.

2. Quantile inequalities

2.1. The KMT quantile inequality

The following quantile inequality is essentially due to KMT [31] and it can be implied from their analysis. That it holds more generally than in the i.i.d. sum setup of Proposition [KMT] is more or less known. (See Remark 1 below.) The proof that we provide here basically follows the lines of that given for the special case of the standardized Binomial random variable (8) with parameters $p = 1/2$ and n in Section 1.2 of Mason [37]. This proof, in turn, was largely adapted from notes taken from the *Diplomarbeit* of Richter [44]. Very similar details are to be found in Einmahl [18]. Let $\{F_n\}_{n\geq 1}$ be a sequence of cdfs, not necessarily being that of a sequence of sums of i.i.d. random variables of the form (5), and let Y_n be defined as in (4).

Theorem 1. *With the above notation, assume there exist a sequence $K_n > 0$, a sequence $0 < \varepsilon_n < 1$ and an integer $n_0 \geq 1$ such that for all $n \geq n_0$ and $0 < z \leq \varepsilon_n \sqrt{n}$*

$$P\left\{Y_n > z\right\} \leq \left(1 - \Phi\left(z\right)\right) \exp\left(K_n\left(z^3 + 1\right)/\sqrt{n}\right), \tag{10}$$

$$P\left\{Y_n > z\right\} \geq \left(1 - \Phi\left(z\right)\right) \exp\left(-K_n\left(z^3 + 1\right)/\sqrt{n}\right), \tag{11}$$

$$P\left\{Y_n < -z\right\} \leq \Phi\left(-z\right) \exp\left(K_n\left(z^3 + 1\right)/\sqrt{n}\right), \tag{12}$$

and

$$P\left\{Y_n < -z\right\} \geq \Phi\left(-z\right) \exp\left(-K_n\left(z^3 + 1\right)/\sqrt{n}\right). \tag{13}$$

Then whenever $n \geq n_0 \vee \left(64K_n^2\right)$ *and*

$$|Y_n| \leq \eta_n \sqrt{n}, \tag{14}$$

where $\eta_n = \varepsilon_n \wedge \left(1/\left(8K_n\right)\right)$, we have

$$|Y_n - Z| \leq \frac{2K_n Y_n^2}{\sqrt{n}} + \frac{2K_n}{\sqrt{n}}. \tag{15}$$

Remark 1. Though not explicitly stated in KMT [31], Theorem 1 has long been known in one form or another by practitioners in strong approximation theory.

Theorem A in KMT [31] implies that if $X_1, X_2, \ldots,$ are i.i.d. X satisfying the assumptions of Theorem [PS], then the sequence of random variables $\{Y_n\}_{n\geq 1}$ as defined in (5) satisfies (10), (11), (12) and (13). KMT [31] do not provide a proof of their Theorem A, however they refer the reader to a large deviation theorem in Petrov [42] (see Theorem 1 on page 218 of Petrov [42] or Theorem 5.23 of Petrov [43]), where it is pointed out that, though the result is formulated under the more restrictive assumption that $\varepsilon_n \to 0$ as $n \to \infty$, his proof is applicable to establishing this refinement. Direct proofs of the fact that $\{Y_n\}_{n\geq 1}$ fulfills the assumptions of Theorem 1 are given by Einmahl [18] (see his Corollary 1 for a more general result from which this result follows) and Theorem 3.1 of Arak and Zaitsev [2]. Theorem 1 of Einmahl [18] provides conditions under which the assumptions hold for independent but not necessarily identically distributed sums. For further quantile inequalities along this line consult Sakhanenko [46, 48, 49].

Theorem [PS] implies that under its assumptions and on its probability space, as $n \to \infty$,

$$\sigma^{-1}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Z_i + R_n, \text{ a.s.} \tag{16}$$

with rate $R_n = O(\log n)$. We should point out that the improved quantile inequalities that in subsection 2.2 are shown to hold under additional assumptions on X do not in general lead to corresponding improvements in the rate in (16), namely, $O(\log n)$ cannot be replaced by $o(\log n)$. A result of Bártfai [3] implies that this can only happen when $\sigma^{-1}X =_d Z$. Refer to Theorem 2.3.2 in M. Csörgő and Révész [12] and especially to the very nice expository paper on strong invariance principles by P. Major [35]. In short, unless $\sigma^{-1}X =_d Z$, the best rate possible in (16) is $O(\log n)$.

Remark 2. Obviously, whenever $K_n^2/n \to 0$ then there exists an integer $n_1 \geq 1$ such that for all $n \geq n_1$ we have $n \geq n_0 \vee \left(64K_n^2\right)$.

Remark 3. In typical applications $K_n = K$, $\varepsilon_n = \varepsilon$ and $\eta_n = \eta$ for all $n \geq 1$, where K, ε and η are fixed positive constants.

Here is a special case of Theorem 1 that will lead to some interesting applications.

Theorem 2. *Assume there exist an $L > 0$, an $0 < \varepsilon < 1$, a $p \geq 2$ and an integer $n_0 \geq 1$ such that for all $n \geq n_0$ and $0 < z \leq \varepsilon n^{1/p}$*

$$P\left\{Y_n > z\right\} \leq \left(1 - \Phi\left(z\right)\right) \exp\left(L\left(z^3 + 1\right)/n^{1/p}\right), \tag{17}$$

$$P\left\{Y_n > z\right\} \geq \left(1 - \Phi\left(z\right)\right) \exp\left(-L\left(z^3 + 1\right)/n^{1/p}\right), \tag{18}$$

$$P\left\{Y_n < -z\right\} \leq \Phi\left(-z\right) \exp\left(L\left(z^3 + 1\right)/n^{1/p}\right), \tag{19}$$

and

$$P\left\{Y_n < -z\right\} \geq \Phi\left(-z\right) \exp\left(-L\left(z^3 + 1\right)/n^{1/p}\right). \tag{20}$$

Then whenever $n \geq n_0 \vee \left(64L^2 n^{1-2/p}\right)$ *and*

$$|Y_n| \leq \eta n^{1/p}, \tag{21}$$

where $\eta = \varepsilon \wedge (1/(8L))$, *we have*

$$|Y_n - Z| \leq \frac{2LY_n^2}{n^{1/p}} + \frac{2L}{n^{1/p}}. \tag{22}$$

Proof. The proof follows from Theorem 1 by setting $K_n = Ln^{1/2-1/p}$ and $\varepsilon_n = \varepsilon n^{-1/2+1/p}$. □

From Theorem 2 we get the following distributional bound for the coupling $|Y_n - Z|$.

Corollary 1. *In addition to the assumptions of the Theorem 2, assume that for suitable positive constants* a, b *and* c *for all* $n \geq 1$ *and* $z \geq 0$

$$P\left\{|Y_n| \geq z\right\} \leq c \exp\left(-\frac{bz^2}{1 + a\left(n^{-1/p}z\right)^{2p/(p+2)}}\right). \tag{23}$$

Then for positive constants C *and* λ, *for all* $z \geq 0$ *and* $n \geq 1$,

$$P\left\{n^{1/p}|Y_n - Z| > z\right\} \leq C \exp\left(-\lambda z^{4/(p+2)}\right), \tag{24}$$

where Y_n *is defined as in (4).*

For conditions that imply that the assumptions of Corollary 1 hold refer to Prob-example 1 in subsection 3.1.

Here is an interesting application of Theorem 1 and the methods of proof of Corollary 1 to martingale difference sequences. It shows how to apply Theorem 1 when the parameters depend on n.

Corollary 2. *Let* $(\xi_i, \mathcal{F}_i)_{i=0,\ldots,n}$ *be a square integrable martingale difference sequence with* $\xi_0 = 0$, *and* $\mathcal{F}_0 \subset \cdots \subset \mathcal{F}_n \subset \mathcal{F}$ *for* $n = 1, 2, \ldots$, *satisfying*

$$\max_i |\xi_i| \leq L \tag{25}$$

and

$$\left|\sum_{i=1}^n E(\xi_i^2 | \mathcal{F}_{i-1}) - n\right| \leq M^2, \tag{26}$$

where L *and* M *are finite positive constants. Also assume that for all* $n \geq 1$

$$\sum_{i=1}^n E\xi_i^2 = n. \tag{27}$$

Then there exist constants $\alpha > 0$ *and* $D > 0$ *and an integer* $n_1 \geq 2$ *such that whenever* $n \geq n_1$ *and*

$$|Y_n| \leq \alpha \sqrt[4]{n}, \tag{28}$$

where $Y_n =_d \sum_{i=1}^{n} \xi_i/\sqrt{n}$ *and is defined as in (4), we have*

$$\sqrt{n}\,|Y_n - Z|\,/\log n \le 2D\left(Y_n^2 + 1\right). \tag{29}$$

Furthermore, there exist positive constants C *and* λ *such that for all* $z \ge 0$ *and* $n \ge 1$,

$$P\left\{\sqrt[4]{n}\,|Y_n - Z| > z\right\} \le C\exp\left(-\lambda z\right). \tag{30}$$

By repeated application of Lemma A.1 of Berkes and Philipp [4] in combination with the Kolmogorov extension theorem we obtain the following proposition.

Proposition 1. *Let* $\xi_1, \xi_2, \ldots,$ *be a sequence of random variables on the same probability space. For each integer* $n \ge 1$ *let* g_n *be a measurable function from* $\mathbb{R}^d$ *to* $\mathbb{R}$ *and* p_n *be a non-negative function defined on* $[0,\infty)$. *Suppose that for each integer* $n \ge 1$ *there exists a probability space on which sit* $(\widetilde{\xi}_1, \ldots, \widetilde{\xi}_n)$ *such that* $(\widetilde{\xi}_1, \ldots, \widetilde{\xi}_n) =_d (\xi_1, \ldots, \xi_n)$ *and a standard normal random variable* $\widetilde{Z}_n$ *such that for all* $z \ge 0$ *and* $n \ge 1$,

$$P\left\{\left|g_n\left(\widetilde{\xi}_1, \ldots, \widetilde{\xi}_n\right) - \widetilde{Z}_n\right| > z\right\} \le p_n(z). \tag{31}$$

Then one can construct a probability space on which sit a sequence of random variables $\widetilde{\xi}_1, \widetilde{\xi}_2, \ldots$ *having the same distribution as* $\xi_1, \xi_2, \ldots,$ *and a sequence of standard normal random variables* $\widetilde{Z}_1, \widetilde{Z}_2, \ldots,$ *such that inequality (31) holds for each* $n \ge 1$.

Here is an example of the use of Proposition 1.

Corollary 3. *Let* $X, X_1, X_2, \ldots$ *be a sequence of i.i.d. mean zero random variables with variance* $0 < \sigma^2 < \infty$, *such that* $E\exp(a|X|) < \infty$ *for some* $a > 0$. *Further for each* $n \ge 1$, *let* $\sigma_{1.n}, \ldots, \sigma_{n.n}$ *be an array of constants satisfying (i)* $\sum_{i=1}^{n} \sigma_{i.n}^2 = 1$ *and (ii) for some* $c > 0$, $\max_{1 \le i \le n} |\sigma_{i.n}| \le c/\sqrt{n}$. *Then i.i.d.* $X, X_1, X_2, \ldots$ *random variables and a sequence of standard normal random variables* $Z_1, Z_2, \ldots$ *can be put on the same probability space so that for suitable constants* $C > 0$ *and* $\lambda > 0$ *one has for all* $x \ge 0$

$$P\left(\sqrt{n}\left|\sigma^{-1}\sum_{i=1}^{n}\sigma_{i.n}X_i - Z_n\right| > x\right) \le C\exp\left(-\lambda x\right).$$

This can be shown by using Corollary 1 with $p = 2$, in combination with Proposition 1. In particular, that inequality (23) holds for appropriate constants a, b and c follows from an application of the classic Bernstein inequality (cf. p. 855 of Shorack and Wellner [53]). To verify that the conditions of Theorem 2 are satisfied with $p = 2$ we apply Corollary 1 in Einmahl [18].

2.1.1. Some remarks on strong approximations

Let $\mathcal{X} = \{X_n,\ n \ge 1\}$ denote a sequence of independent mean zero random variables such that X_n has variance $0 < \sigma_n^2 < \infty$, $n \ge 1$, and let $\mathcal{Y} = \{Y_n,\ n \ge 1\}$

be a sequence of independent mean zero normal random variables such that each Y_n has variance $0 < \sigma_n^2 < \infty$, $n \geq 1$. Whenever $\mathcal{X}$ and $\mathcal{Y}$ are on the same probability space set for each $n \geq 1$

$$\Delta_n(\mathcal{X}, \mathcal{Y}) = \max_{1 \leq j \leq n} \left| \sum_{i=1}^{j} X_i - \sum_{i=1}^{j} Y_i \right|.$$

Consider the special case when $\mathcal{X}$ is an i.i.d. X sequence of random variables, where $EX = 0$ and $0 < VarX = \sigma^2 < \infty$. Theorem [PS], i.e. Theorem 1 of KMT [32], implies that whenever, in addition, $E\exp(a|X|) < \infty$ for some $a > 0$, then on the same probability space one can define $\mathcal{X}$ and $\mathcal{Y}$ such that

$$\Delta_n(\mathcal{X}, \mathcal{Y}) = O(\log n), \text{ a.s.} \tag{32}$$

If, instead, we assume that $E|X|^r < \infty$ for some $r > 2$, we have

$$\Delta_n(\mathcal{X}, \mathcal{Y}) = o\left(n^{1/r}\right), \text{ a.s.} \tag{33}$$

The case $2 < r \leq 3$ is the Corollary in Major [34] and the case $r > 3$ is Theorem 2 of KMT [32].

Couplings such that almost sure statements as (32) and (33) hold allow one to infer laws of the iterated logarithm [LILs] for the i.i.d. partial sums $\sum_{i=1}^{n} X_i$ from the LIL for the partial sums of i.i.d. normal random variables $\sum_{i=1}^{n} Y_i$. These are special cases of *strong approximations.* To learn more about strong approximations and their applications refer to M. Csörgő and Révész [12].

Corollary 3 is not a strong approximation. However, Proposition 1, upon which it is based, and versions of it play a vital role in establishing strong approximations via quantile and conditional quantile coupling inequalities, combined with dyadic construction schemes and blocking arguments.

The classical KMT approximation results for i.i.d. sums have been extended by Sakhanenko to the independent but not necessarily i.d. case. Here are his two main coupling inequalities. Quantile and conditional quantile inequalities play an indispensable role in establishing these results.

Theorem A (Sakhanenko [46])**.** *Let $\mathcal{X} = \{X_n,\ n \geq 1\}$ be a sequence of independent mean zero random variables such that X_n has variance $0 < \sigma_n^2 < \infty$, $n \geq 1$. Suppose there exists a $\lambda > 0$ such that for all $n \geq 1$*

$$\lambda E|X_n|^3 \exp(|X_n|) \leq EX_n^2. \tag{34}$$

Then on the same probability space one can define $\mathcal{X}$ and $\mathcal{Y}$ such that for all $n \geq 1$

$$E\exp(\lambda A \Delta_n(\mathcal{X}, \mathcal{Y})) \leq 1 + \lambda \sum_{i=1}^{n} EX_i^2, \tag{35}$$

where A is a universal constant.

Theorem B (Sakhanenko [48])**.** *Let $\mathcal{X} = \{X_n,\ n \geq 1\}$ be a sequence of independent mean zero random variables such that X_n has variance $0 < \sigma_n^2 < \infty$, $n \geq 1$. Suppose that for some $r > 2$, $E|X_n|^r < \infty$ for all $n \geq 1$. Then on the same probability space one can define $\mathcal{X}$ and $\mathcal{Y}$ such that for all $n \geq 1$*

$$E\left|\Delta_n\left(X,Y\right)\right|^r \leq \left(Cr\right)^r \sum_{i=1}^{n} E\left|X_i\right|^r, \tag{36}$$

where C is a universal constant.

Theorems A and B were announced in Sakhanenko [45]. Theorem A is proved in Sakhanenko [46], where it is appears as Theorem 1, and Theorem B is established in Sakhanenko [48], where it is stated and proved in Section 5 of his paper as Corollary 5. (An earlier version with worse constants is given in Sakhanenko [47].) The above formulations of the Sakhanenko coupling inequalities were adapted from those given in Shao [52]. (Actually our Theorem B implies Shao's version via Markov's inequality.) He uses Theorems A and B to establish strong approximations for partial sums of independent random variables, and shows how they lead to LILs and strong laws.

2.2. Refined quantile inequalities

To formulate the results in this section we shall need the following regularity condition. Let $\{Y_n\}_{n\geq 1}$ be a sequence of random variables, where each Y_n has cdf F_n. Assume that for every $n \geq 1$ there is an $\varepsilon_n > 0$ such that

$$0 < F_n\left(-\varepsilon_n-\right) \leq F_n\left(\varepsilon_n\right) < 1. \tag{37}$$

Note that this assumption is very weak. It holds as long as $EY_n = 0$ and $\mathbb{P}\left(Y_n = 0\right) < 1$, for instance, when for each $n \geq 1$, Y_n is as in (5), with $X_1, X_2, \ldots$, i.i.d. X, where X is nondegenerate having expectation zero.

We shall show that the KMT quantile coupling for Y_n as in (5) given in Proposition [KMT] can be improved with a rate $1/\sqrt{n}$, when X satisfies additional regularity conditions. Our main result in this section is the following refined KMT quantile coupling inequality.

Theorem 3. *In addition to the assumptions of Proposition [KMT] suppose that $EX^3 = 0$ and its characteristic function $v(t)$ satisfies*

$$\limsup_{|t|\to\infty} |v(t)| < 1. \tag{38}$$

Then there exist $C > 0$ and $\varepsilon > 0$ such that for every integer $n \geq 1$, with Y_n defined as in (5) and (4),

$$|Y_n - Z| \leq \frac{C}{n} + \frac{C}{n}|Y_n|^3$$

for $|Y_n| \leq \varepsilon\sqrt{n}$.

Notice that if the random variable X is symmetric then $EX^3 = 0$. Also if the absolutely continuous component of the distribution of the random variable X is nonzero, one can readily conclude by the Riemann–Lebesgue lemma that assumption (38) is satisfied. Theorem 3 will be an immediate consequence of Theorem 4 and Proposition 2 below.

Our next result discloses a relationship between the existence of a certain type of large deviation result and a sharp quantile coupling inequality. Such a large deviation is often called a "Petrov expansion". Actually, the expansion that we use in this paper is even more "precise" than that of Petrov (see Remark 4), and perhaps it is better to call it a Saulis expansion (see pages 249 and 169 in Petrov [42]).

Note in this paper, we use the notation $A_n(x) = O(a_n(x))$ with $x \in D_n$, where $\{D_n\}_{n\geq 1}$ is a sequence of sets, to mean that there is an $n_0 \geq 1$ and $C > 0$ such that for all $n \geq n_0$

$$-Ca_n(x) \leq A_n(x) \leq Ca_n(x)$$

uniformly over all $x \in D_n$.

Theorem 4. *Let Z be a standard normal random variable. Let Y_n be a sequence of random variables. Assume that there is a positive ε such that*

$$P\{Y_n > x\} = \overline{\Phi}(x) \exp\left(O\left(n^{-1}x^4 + n^{-1}\right)\right),$$
$$P\{Y_n < -x\} = \Phi(-x) \exp\left(O\left(n^{-1}x^4 + n^{-1}\right)\right),$$

where $\overline{\Phi}(x) = 1 - \Phi(x)$, $O\left(n^{-1}x^4 + n^{-1}\right)$ is uniform on the interval $x \in [0, \varepsilon\sqrt{n}]$, and (37) holds. Then there exist $C_1 > 0$ and $\varepsilon_1 > 0$ such that for every $n \geq 1$, with Y_n defined as in (4),

$$|Y_n - Z| \leq \frac{C_1}{n} + \frac{C_1}{n}|Y_n|^3 \tag{39}$$

for $|Y_n| \leq \varepsilon_1\sqrt{n}$.

Remark 4. Let

$$a(n,x) = n^{-1/2}x^3 + n^{-1/2}x + n^{-1/2}.$$

The Petrov expansion is obtained by replacing the $O\left(n^{-1}x^4 + n^{-1}\right)$ in Theorem 4 by $O(a(n,x))$ (see Theorem 1 in Chapter VIII of Petrov [42], or Theorem A in Komlós, Major, and Tusnády [31]). In this case, as in Theorem 1 above, the corresponding coupling inequality becomes

$$|Y_n - Z| \leq \frac{C_1}{\sqrt{n}} + \frac{C_1}{\sqrt{n}}|Y_n|^2$$

(see Sakhanenko [46, 49]). The deviation term $O\left(n^{-1}x^4 + n^{-1}\right)$ improves the $O(a(n,x))$ term with a rate $1/\sqrt{n}$ uniformly in $x \in [0, a]$ for any $a > 0$, which shows in the corresponding quantile coupling inequality.

In some applications, it is more convenient to use the following corollary, where the bound involves only the standard normal random variable Z. Zhou [57] used such a coupling of a standardized Beta random variable with a standard normal random to establish asymptotic equivalence of Gaussian variance regression and Gaussian white noise with a drift. He found that his analysis was much easier when he used the following bound in his moment calculations.

Corollary 4. *Under the assumptions of Theorem 4, there exist $C > 0$ and $\varepsilon > 0$ such that for every $n \geq 1$*

$$|Y_n - Z| \leq \frac{C}{n}\left(1 + |Z|^3\right), \text{ whenever } |Z| \leq \varepsilon\sqrt{n}, \tag{40}$$

where Y_n is defined as in (4).

We have the following Saulis expansion (see page 188 in Saulis and Statulevicius [51]), which shows that the conditions of Theorem 4 hold when those of Theorem 3 are satisfied. Note that the following proposition, when combined with Theorem 4 establishes Theorem 3.

Proposition 2. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. X random variables for which $EX = 0$, $EX^2 = 1$, $EX^3 = 0$ and $E\exp(a|X|) < \infty$ for some $a > 0$. Further assume that (38) holds. Then for Y_n is as in (5) and (4), there exists a positive constant η such that*

$$P\{Y_n > x\} = \overline{\Phi}(x)\exp\left(O\left(n^{-1}x^4 + n^{-1}\right)\right), \tag{41}$$

$$P\{Y_n < -x\} = \Phi(-x)\exp\left(O\left(n^{-1}x^4 + n^{-1}\right)\right), \tag{42}$$

in the interval $0 \leq x \leq \eta\sqrt{n}$.

Proof. We only verify (42). The proof that (41) holds is similar. Saulis [50] shows that there is a constant $\eta > 0$ such that for some $C > 0$ and n sufficiently large,

$$P\{Y_n < -x\} \leq \Phi(-x)\exp\left(Cn^{-1}x^4 + Cn^{-1}\right)$$

and

$$P\{Y_n < -x\} \geq \Phi(-x)\exp\left(-Cn^{-1}x^4 - Cn^{-1}\right)$$

for $1 \leq x \leq \eta\sqrt{n}$, when the third moment of X is 0. See also page 249 of Petrov [42].

Theorem 3 in page 169 of Petrov [42] together with $EX^3 = 0$ imply

$$|P\{Y_n < -x\} - \Phi(-x)| = O\left(\frac{1}{n}\right)$$

uniformly over $0 \leq x \leq 1$, i.e., $P(Y_n < -x) = \Phi(-x)\exp\left(O\left(n^{-1}\right)\right)$. □

The proof of Proposition 2 can also be derived by arguments similar to those in Section 8.2 of Petrov [42]. Also note that the above expansion holds when the random variables X_i are replaced by $-X_i$. This implies that the expansion above holds when "$<$" is replaced by "$\leq$".

Other couplings are possible. Consider the following coupling result, which can yield refinements when for each $n \geq 1$ the distribution of Y_n is concentrated on a lattice.

Theorem 5. *Let Z be a standard normal random variable. Let Y_n be a sequence of random variables. Assume that there exists a sequence of sets $\{\mathbb{C}_n\}_{n\geq 1}$ such that $P\{Y_n \in \mathbb{C}_n\} = 1$ and a positive ε such that for all $n \geq 1$,*

$$P\{Y_n > x\} = \overline{\Phi}(x) \exp\left(O\left(n^{-1}x^4 + n^{-1/2}\right)\right),$$
$$P\{Y_n < -x\} = \Phi(-x) \exp\left(O\left(n^{-1}x^4 + n^{-1/2}\right)\right),$$

where $\overline{\Phi}(x) = 1 - \Phi(x)$, $O(n^{-1}x^4 + n^{-1/2})$ is uniform in the set $[0, \varepsilon\sqrt{n}] \cap \mathbb{C}_n$, and (37) holds. Moreover, the expansions above hold when "$<$" is replaced by "$\leq$". Then there exist $C_1 > 0$ and $\varepsilon_1 > 0$ such that for every $n \geq 1$, with Y_n defined as in (4),

$$|Y_n - Z| \leq \frac{C_1}{\sqrt{n}} + \frac{C_1}{n}|Y_n|^3 \tag{43}$$

for $|Y_n| \leq \varepsilon_1\sqrt{n}$.

Results in Carter and Pollard [9] show that the assumptions of Theorem 5 hold when Y_n is the standardized sum of i.i.d. Bernoulli(1/2) with

$$\mathbb{C}_n = \left\{2\left(x - n/2\right)/\sqrt{n} : x = 0, 1, \ldots, n\right\}. \tag{44}$$

Arguing as in the proof of Corollary 4, we obtain:

Corollary 5. *Under the assumptions of Theorem 5, there exist $C > 0$ and $\varepsilon > 0$ such that for every $n \geq 1$, with Y_n defined as in (4),*

$$|Y_n - Z| \leq \frac{C}{\sqrt{n}} + \frac{C}{n}|Z|^3, \text{ whenever } |Z| \leq \varepsilon\sqrt{n}.$$

Remark 5. If the sequence of random variables $\{Y_n\}_{n\geq 1}$ in Theorems 4 or 5 does not satisfy condition (37), then their conclusions hold for all large enough n.

To see what kind of couplings one can get for standardized partial sums of i.i.d. X random variables when the condition that $E\exp(a|X|) < \infty$ for some $a > 0$ is replaced by the assumption that $E\exp\{g(|X|)\} < \infty$ for a suitable continuous increasing function g on $[0, \infty)$, refer to Appendix C.

3. Applications of quantile couplings

3.1. Applications to probability theory

Prob-example 1 (Partial sums) Assume $X, X_1, \ldots, X_n$ are i.i.d. with $EX = 0$, $0 < VarX = \sigma^2 < \infty$, satisfying for some $\gamma \geq 0$ and $K \geq \sigma$

$$E|X^m| \leq (m!)^{1+\gamma} K^{m-2}\sigma^2, \text{ for } m \geq 3,$$

or in terms of for $p \geq 2$,

$$E|X^m| \leq (m!)^{(p+2)/4} K^{m-2}\sigma^2, \text{ for } m \geq 3. \tag{45}$$

(Note that the case $p = 2$ is the classic Bernstein condition.) An application of Theorem 3.1 in Saulis and Statulevičius [51] shows that the sequence of random variables

$$Y_n =_d \sum_{i=1}^{n} X_i / \left(\sigma\sqrt{n}\right),$$

satisfies the conditions of Corollary 1. The case $p = 2$, i.e. $\gamma = 0$, corresponds to the conditions of Theorem [PS]. In fact, a random variable X having mean zero and variance $0 < \sigma^2 < \infty$ satisfies (45) if and only if for some $C > 0$ and $d > 0$ for all $x \geq 0$

$$P\{|X| > x\} \leq C\exp\left(-dx^{\beta}\right), \tag{46}$$

where $\beta = 4/(p+2)$. Refer to Appendix B for a proof.

If one also assumes that the characteristic function of X satisfies (38) then one can apply an expansion due to Wolf [56] to show that the coupling (24) can be improved to say that for positive constants C and λ, for all $z \geq 0$ and $n \geq 1$,

$$P\left\{n^{1/2}|Y_n - Z| > z\right\} \leq C\exp\left(-\lambda z^{4/(p+2)}\right). \tag{47}$$

For a sketch of how this done see Appendix C.

Note that in the following Prob-examples 2-5, it is understood that $K_n = K$, $\varepsilon_n = \varepsilon$ and $\eta_n = \eta$ for all $n \geq 1$, where K, ε and η are fixed positive constants.

Prob-example 2 (Self-normalized sums) Let $X_1, X_2, \ldots,$ be i.i.d. X, where X has mean 0, variance $0 < \sigma^2 < \infty$ and finite third absolute moment $E|X|^3 < \infty$. For each integer $n \geq 1$ consider the self-normalized sum

$$Y_n =_d \frac{X_1 + \cdots + X_n}{\sqrt{X_1^2 + \cdots + X_n^2}}.$$

A special case of the results of Jing, Shao and Wang [29] (also see Theorem 7.1 in de la Peña, Lai and Shao [14]) shows that for $0 \leq x \leq \sqrt{n}$,

$$\frac{P\{Y_n > x\}}{1 - \Phi(x)} = \exp\left(O(1)\,\Delta_{n,x}\right) \text{ and } \frac{P\{Y_n < -x\}}{\Phi(-x)} = \exp\left(O(1)\,\Delta_{n,x}\right), \tag{48}$$

where

$$\Delta_{n,x} = \frac{(1+x)^2 E|X|^2 1\{|X| > \sigma\sqrt{n}/(1+x)\}}{\sigma^2} + \frac{(1+x)^3 E|X|^3 1\{|X| \leq \sigma\sqrt{n}/(1+x)\}}{\sigma^3\sqrt{n}}.$$

Clearly

$$\frac{(1+x)^3 E|X|^3 1\{|X| \le \sigma\sqrt{n}/(1+x)\}}{\sigma^3\sqrt{n}} \le \frac{(1+x)^3 E|X|^3}{\sigma^3\sqrt{n}}.$$

Note that by Hölder's inequality

$$E|X|^2 1\left\{|X| > \sigma\sqrt{n}/(1+x)\right\} \le \left(E|X|^3\right)^{2/3} \left(P\left\{|X| > \sigma\sqrt{n}/(1+x)\right\}\right)^{1/3},$$

which by Markov's inequality is

$$\le \frac{(1+x)E|X|^3}{\sigma\sqrt{n}}.$$

Thus

$$\frac{(1+x)^2 E|X|^2 1\{|X| > \sigma\sqrt{n}/(1+x)\}}{\sigma^2} \le \frac{(1+x)^3 E|X|^3}{\sigma^3\sqrt{n}}.$$

Hence

$$\Delta_{n,x} \le \frac{2(1+x)^3 E|X|^3}{\sigma^3\sqrt{n}} \le \frac{8\left(1+x^3\right)E|X|^3}{\sigma^3\sqrt{n}}.$$

Therefore by (48), conditions (10), (11), (12) and (13) hold for a suitable $K > 0$ and all $n \ge n_0$ for some integer $n_0 \ge 1$ and with $\varepsilon = 1$. Hence by Theorem 1 the quantile inequality (15) holds whenever $n \ge n_0 \vee \left(64K^2\right)$ and (14) is satisfied with $\eta = 1 \wedge (1/(8K))$.

Now since (obviously) $Y_n = O_p(1)$, we can apply Theorem 2.5 of Giné, Goetze and Mason [20] to show that for suitable constants $b > 0$ and $c > 0$, for all $z \ge 0$ and $n \ge 1$

$$P\{|Y_n| \ge z\} \le c\exp\left(-bz^2\right). \tag{49}$$

Thus we conclude by Corollary 1 that (24) with $p = 2$ holds in this example.

Recall in Prob-example 1 that for inequality (24) with $p = 2$ to hold for an unself-normalized sum $\sigma^{-1}S_n$ we required that X have a finite moment generating function in a neighborhood of zero. Prob-example 2 shows that self-normalizing dramatically reduces the assumptions needed for (24) with $p = 2$ to be valid.

Prob-example 3 (Finite sampling without replacement) Let

$$\{a_{1,m}, \dots, a_{m,m},\ m \ge 1\}$$

be a triangular array of constants. For each integer $m \ge 1$ let

$$\mu_m = \frac{a_{1,m} + \cdots + a_{m,m}}{m},\ \sigma_m^2 = \frac{1}{m}\sum_{i=1}^{m}(a_{i,m} - \mu_m)^2$$

and

$$\beta_{3,m} = \frac{1}{m}\sum_{i=1}^{m}\left|a_{i,m}-\mu_m\right|^3/\sigma_m^3.$$

For each $m \geq 1$ choose $1 \leq n_m \leq m$ and set

$$p_m = \frac{n_m}{m},\ q_m = 1-\frac{n_m}{m} \text{ and } \omega_m^2 = mp_mq_m.$$

Assume that for all $m \geq m_0$ for some integer $m_0 \geq 1$,

$$0 < \alpha_1 \leq M_m := \max_{1\leq i\leq m}|a_{i,m}-\mu_m| \leq \alpha_2 < \infty,\ \text{for } 0 < \alpha_1 \leq \alpha_2 < \infty; \tag{50}$$

$$0 < \beta_1 \leq \beta_{3,m} \leq \beta_2 < \infty,\ \text{for } 0 < \beta_1 \leq \beta_2 < \infty; \tag{51}$$

$$0 < \tau_1 \leq \ \sigma_m \leq \tau_2 < \infty, \text{ for } 0 < \tau_1 \leq \tau_2 < \infty; \tag{52}$$

and

$$0 < \rho_1 m \leq mp_mq_m \leq \rho_2 m, \text{ for } 0 < \rho_1 \leq \rho_2 < \infty. \tag{53}$$

For each integer $m \geq 1$ let S_m denote the sum of $X_1, \ldots, X_{n_m}$ taken by simple random sampling without replacement from $\{a_{1,m}, \ldots, a_{m,m}\}$. Let

$$Y_m =_d \frac{S_m - \mu_m}{\sigma_m\omega_m}.$$

Assumptions (50), (51), (52) and (53) permit us to apply Theorem 1.1 of Hu, Robinson and Wang [28] to get for a suitable constant $K > 0$ and $\varepsilon > 0$ that (10), (11), (12) and (13) hold for all $m \geq m_0$. (For an earlier version of their result where $m = 2n$ and $n_m = n$, $n \geq 1$, refer to Lemma 3 of KMT [32].) Therefore we can apply Theorem 1 (note we replace n by m) to show that the quantile inequality (15) is valid whenever $m \geq m_0 \vee \left(64K^2\right)$ and (14) is satisfied with $\eta = 1 \wedge (1/(8K))$.

Notice that for all $m \geq m_0$ for some $\lambda > 0$

$$\frac{M_m}{\sigma_m\omega_m} \leq \frac{1}{\lambda\sqrt{m}}.$$

This bound combined with the Hoeffding [27] inequality for simple random sampling from a finite population without replacement gives for all $z \geq 0$ and $m \geq m_0$

$$P\left\{|Y_m| \geq z\right\} \leq 2\exp\left(-\frac{\lambda^2z^2}{2}\right). \tag{54}$$

This allows us to apply Corollary 1 to conclude that (24) with $p = 2$ holds for all $m \geq m_0$.

A more carefully analysis leads to an inequality of the form: For all $z \geq 0$ and $m \geq m_0$,

$$P\left\{\sqrt{m}\,|Y_m - Z| > z\right\} \leq C_m\exp\left(-\lambda_m z\right), \tag{55}$$

where C_m and λ_m are positive constants depending on m, ω_m, σ_m, M_m and $\beta_{3,m}$. For a closely related result in the special case when $a_{i,m} \in \{-1, 1\}$ for $i = 1, \dots, m$, $m \geq 1$, refer to Theorem 3.2 of Chatterjee [10].

Prob-example 4 Let $X_1, X_2, \dots,$ be a stationary sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, satisfying $EX_1 = 0$, $VarX_1 = 1$. Set $S_n = X_1 + \cdots + X_n$ and $B_n^2 = Var(S_n)$. Assume that for some $\sigma_0^2 > 0$ we have $B_n^2 \geq \sigma_0^2 n$ for all $n \geq 1$. Set $Y_n =_d S_n / B_n$.

Let $\{\mathcal{F}_s^t : 1 \leq s \leq t < \infty\}$ be a family of $\sigma-$algebras such that (i) $\mathcal{F}_s^t \subset \mathcal{F}$ for all $1 \leq s \leq t < \infty$, (ii) $\mathcal{F}_{s_1}^{t_1} \subset \mathcal{F}_{s_2}^{t_2}$ for all $1 \leq s_2 \leq s_1 \leq t_1 \leq t_2 < \infty$, (iii) $\sigma\{X_u, 1 \leq s \leq u \leq t < \infty\} \subset \mathcal{F}_s^t$.

Define the $\alpha-$mixing, $\varphi-$mixing and $\psi-$mixing functions

$$\alpha(s,t) = \sup_{A \in \mathcal{F}_1^s, B \in \mathcal{F}_t^\infty} |P(A \cap B) - P(A)P(B)|,$$

$$\varphi(s,t) = \sup_{A \in \mathcal{F}_1^s, B \in \mathcal{F}_t^\infty} \left| \frac{P(A \cap B) - P(A)P(B)}{P(A)} \right|,$$

$$\psi(s,t) = \sup_{A \in \mathcal{F}_1^s, B \in \mathcal{F}_t^\infty} \left| \frac{P(A \cap B) - P(A)P(B)}{P(A)P(B)} \right|.$$

In these last two expressions it is understood that $0/0 := 0$, whenever its occurs. Introduce the mixing rates: For some $M > 0$ and $\mu > 0$

(M.1) $\alpha(s,t) \leq Me^{-\mu(t-s)}$, (M.2) $\varphi(s,t) \leq Me^{-\mu(t-s)}$, (M.3) $\psi(s,t) \leq Me^{-\mu(t-s)}$,

and the bounding conditions:

(B.1) $|X_1| \leq C$ for some $0 < C < \infty$, (B.2) $Ee^{\theta|X_1|} < \infty$ for some $\theta > 0$.

Applying results in Statulevičius and Jakimavičius [54] we get that inequalities (23) and (24) hold for suitable positive constants and with p as indicated:

Under conditions (M.1) and (B.1), $p = 6$;
under conditions (M.1) and (B.2), $p = 10$;
under $m-$dependence and (B.1), $p = 2$;
under $m-$dependence and (B.2), $p = 6$;
under conditions (M.2) and (B.1), $p = 2$;
under conditions (M.2) and (B.2), $p = 6$;
under conditions (M.3) and (B.2), $p = 2$.

In the last three cases we assume that the random variables X_t are connected by a Markov chain.

Prob-example 5 (Sample median) Let $\{Y_n\}_{n \geq 1}$ be a sequence of random variables of the form

$$Y_n = 2\sqrt{2n}\left\{V_n - \frac{1}{2}\right\}, \tag{56}$$

where V_n is a $Beta\,(n,n)$ random variable. A long but elementary analysis based on Stirling's formula shows that this sequence satisfies the assumptions of Theorem 1. For more about this example see Exp-example 2 in subsection 3.2. Furthermore, since V_n is equal in distribution to the n^{th} order statistic of $2n-1$ independent Uniform $(0,1)$ random variable, i.e. the *sample median*, we get using identity (11) on page 86 of Shorack and Wellner [53] that for any $z \geq 0$

$$P\left\{|Y_n| > z\right\} \leq P\left\{\sup_{0\leq t\leq 1}\left|\alpha_{2n-1}\left(t\right)\right| > \frac{z}{2}\frac{\sqrt{2n-1}}{\sqrt{2n}}\right\},$$

where $\alpha_{2n-1}\left(t\right)$ is defined as in (7), which by the Dvoretzky-Kiefer-Wolfowitz inequality (see Massart [39]) is

$$\leq 2\exp\left(-z^2\left(2n-1\right)/\left(4n\right)\right) \leq 2\exp\left(-z^2/4\right).$$

Thus we can apply Corollary 1 to get that inequality (24) with $p=2$ is satisfied for all $n \geq 1$. Hence we can conclude that for positive constants C and λ, for all $z \geq 0$ and $n \geq 1$,

$$P\left\{n^{1/2}\left|Y_n - Z\right| > z\right\} \leq C\exp\left(-\lambda z\right), \tag{57}$$

where Y_n is defined as in (56) and (4).

The coupling (57) can be used to give a fairly direct proof of Theorem [PS] in the special case when $X_1, X_2, \ldots,$ are i.i.d. $\omega - 1$, where ω is exponential with mean 1. This exponential special case is important in constructing a Brownian bridge approximation to the uniform quantile process. For details refer to M. Csörgő and Révész [12] and M. Csörgő, S. Csörgő, Horváth and Mason [13].

We conclude this subsection with a remark about probability spaces.

Remark 6. Using Proposition 1 one can construct a probability space on which sit a square integrable martingale difference sequence $(\xi_i, \mathcal{F}_i)_{i=0,\ldots,n}$ satisfying the conditions of Corollary 2 and a sequence of standard normal random variables $\{Z_n\}_{n\geq 1}$ such that for each $n \geq 1$, (30) holds with $Y_n = \sum_{i=1}^{n}\xi_i/\sqrt{n}$ and with Z replaced by Z_n. Analogous statements are true for Prob-examples 1, 2 and 4. By the way, the probability space of Theorem [EP] is constructed in this way.

3.2. Applications to asymptotic equivalence of experiments

Since Donoho and Johnstone [15], a Besov smoothness constraint has been standard in the study of the asymptotic optimally of nonparametric estimation procedures. More recently, under a sharp Besov smoothness assumption and via the Carter and Pollard [9] improved Tusnády inequality, Brown, Carter, Low and Zhang [BCLZ] [7] extended the asymptotic equivalence result of Nussbaum [41] for density estimation. BCLZ [7] is considered to be an important paper in this area. We point out that the Tusnády inequality given in Proposition [T] may not be strong enough to establish their results.

Quantile coupling inequalities of the kind we are discussing in this survey have led to extensions of the asymptotic equivalence theory for density estimation in Nussbaum [41] to general nonparametric estimation models (see Grama and Nussbaum [23, 24, 25]). Among these is the important case of the asymptotic equivalence of Gaussian variance regression to Gaussian white noise.

Zhou [57] and Golubev, Nussbaum and Zhou [21] obtained equivalence results for spectral density estimation. One of their crucial tools is a sharp quantile coupling bound between a standardized Beta random variable and a standard normal random variable to obtain an asymptotic equivalence theory for the Gaussian variance regression**.** This was the key to establish the asymptotic equivalence of spectral density estimation and Gaussian white noise experiments under a Besov smoothness constraint. One of its interesting applications is a coupling of a sample median statistic with a standard normal random variable. It improves upon the classical quantile coupling bounds with a rate $1/\sqrt{n}$, under certain smoothness conditions on the underlying distribution function and it includes the Cauchy distribution as a special case. It is likely that this coupling will be of independent interest because of the fundamental role played by the sample median in statistics.

Here are some more detailed descriptions of these applications.

Exp-example 1 (Asymptotic equivalence of density estimation and Gaussian white noise) Consider the two sequences of experiments:

$$\begin{aligned}
&\mathbb{E}_n : y(1), \dots, y(n), \text{ i.i.d. with density } f \text{ on } [0,1] \\
&\mathbb{F}_n : dy_t = f^{1/2}(t)\, dt + \frac{1}{2} n^{-1/2} dW_t,
\end{aligned}$$

where W_t is a standard Wiener process on $[0,1]$. The asymptotic equivalence of these two sequences was established in BLCZ [7] under a Besov smoothness constraint. The basic approach in their paper is the utilization of the classical KMT [31] construction. To do this they needed the following coupling of a standardized Binomial random variable and a standard normal random variable Z. Let $X_1, X_2, \dots, X_n$ be i.i.d. Bernoulli(1/2). Our Corollary 5 tells us that for every $n \geq 1$ there is a random variable Y_n with

$$Y_n =_d \frac{2\,(X_1 + \cdots + X_n - n/2)}{\sqrt{n}}$$

such that

$$|Y_n - Z| \leq \frac{C}{\sqrt{n}} + \frac{C}{n} |Z|^3$$

for $|Z| \leq \varepsilon\sqrt{n}$, where C, $\varepsilon > 0$ do not depend on n (see also Carter and Pollard [9]). This result was applied in combination with the KMT construction to establish the asymptotic equivalence under a Besov smoothness condition and a compactness in the Besov balls $B^{1/2}_{2,2}$ and $B^{1/2}_{4,4}$ assumption. If instead, one uses the classical Tusnády inequality, a stronger smoothness condition would be needed to establish the asymptotic equivalence.

Exp-example 2 (Asymptotic equivalence of spectral density estimation and Gaussian white noise) Consider the two sequences of experiments:

$$\begin{aligned}&\mathbb{E}_n : y(1), \ldots, y(n), \text{ a stationary centered Gaussian sequence} \\ &\quad \text{with spectral density } f \\ &\mathbb{F}_n : dy_t = \log f(t)\, dt + 2\pi^{1/2} n^{-1/2} dW_t,\end{aligned}$$

where f has support in $[-\pi, \pi]$. The asymptotic equivalence between the sequence of Gaussian spectral density experiments $\mathbb{E}_n$ and the sequence of regression Gaussian white noise experiments $\mathbb{F}_n$ was established in Golubev, Nussbaum and Zhou [21] under a Besov smoothness constraint. In that paper, they used a modification of the dyadic KMT [31]-type construction. Instead of applying a complicated KMT [31] type conditional quantile coupling for higher resolutions, Golubev et al [21] found that in their setup it was easier to use a construction based on the fact that if two X and Y are two independent χ^2_n random variables X and Y then for any $y > 0$,

$$(X | X + Y = y) =_d yB_n,$$

where B_n is a Beta $(n/2, n/2)$ random variable. This permitted them to avoid conditional quantile coupling by considering a coupling for a Beta random variable, obtaining the following coupling inequality, which we get here by an application of Corollary 4. Let Z be a standard normal random variable. For every n, there is a mapping $T_n : \mathbb{R} \mapsto \mathbb{R}$ such that the random variable $B_n = T_n(Z)$ has the Beta $(n/2, n/2)$ law and

$$\left| n\left(B_n - 1/2\right) - \frac{n^{1/2}}{2} Z \right| \le \frac{C}{\sqrt{n}} + \frac{C}{\sqrt{n}} |Z|^3 \tag{58}$$

for $|Z| \le \varepsilon n$, where C, $\varepsilon > 0$ do not depend on n (cf. Zhou [57]).

Exp-example 3 (Quantile coupling of Median statistics) Let $X_1, X_2, \ldots, X_n$ be i.i.d. with density f and let $X_{(1)} \le \cdots \le X_{(n)}$ denote their order statistics. For simplicity, we only consider odd integers $n = 2k + 1$ with $k \ge 0$. Thus, in this notation, the sample median $X_{med} = X_{(k+1)}$. Assume that

$$f(0) > 0,\ f'(0) = 0, \text{ and } f \in C^3$$

and there is an $\epsilon > 0$ such that

$$\int |x|^\epsilon f(x)\, dx < \infty. \tag{59}$$

Let Z be a standard normal random variable. Cai and Zhou [8] show that for every $n \ge 1$, there is a mapping $T_n : \mathbb{R} \mapsto \mathbb{R}$ such that $T_n(Z) =_d X_{med}$ and

$$\left| \sqrt{4n} f(0)\, X_{med} - Z \right| \le C \frac{1}{n} \left(1 + |Z|^3\right), \text{ whenever } |Z| \le \varepsilon\sqrt{n},$$

where C, $\varepsilon > 0$ do not depend on n. For the details of proof and for more general discussions consult Cai and Zhou [8]. It can be shown that our Corollary 4 gives this result. They use this quantile inequality to study the nonparametric location model with Cauchy noise and as well as wavelet regression. Donoho and Yu [16] treated a similar problem, however it is not clear that the minimax property holds for their procedure. In the wavelet regression setting, Hall and Patil [26] studied nonparametric location models and achieved the optimal minimax rate, but under the more restrictive assumption of the existence of a finite fourth moment. Cai and Zhou [8] need only to impose the existence of a finite $\epsilon-$moment (59). The noise can be general and unknown, but yet achieve an optimal minimax rate of convergence. Without the assumptions $f'(0) = 0$ or $f \in C^3$, we can still obtain coupling bounds, but they may not be as tight as the above bound. The tightness of the upper bound affects the underlying smoothness condition that they require in deriving their asymptotic results.

4. Proofs

4.1. The underlying approach to our quantile inequalities

Underlying our quantile inequalites is the following simple observation. Let Y be a random variable with cdf F and as usual let Φ be the cdf of the standard normal random variable Z. Observe trivially that for some y and $u(y) > 0$

$$\Phi(y - u(y)) \leq F(y-) \leq F(y) \leq \Phi(y + u(y)) \tag{60}$$

if and only if

$$y - u(y) \leq \Phi^{-1}(F(y-)) \leq \Phi^{-1}(F(y)) \leq y + u(y). \tag{61}$$

Thus if there is a z such that $y = F^{-1}(\Phi(z))$ we get from (61) and the fact that for all $s \in (0, 1)$

$$F\left(F^{-1}(s)-\right) \leq s \leq F\left(F^{-1}(s)\right)$$

that with $y = F^{-1}(\Phi(z))$ and $s = \Phi(z)$,

$$F^{-1}(\Phi(z)) - u\left(F^{-1}(\Phi(z))\right) \leq z \leq F^{-1}(\Phi(z)) + u\left(F^{-1}(\Phi(z))\right). \tag{62}$$

Now let C be the set of y such that inequality (60) holds and define $Y = F^{-1}(\Phi(Z))$. Clearly whenever $Y \in C$, we have

$$|Y - Z| \leq u(Y).$$

All of our quantile inequalities will follow this approach. For instance, Theorem 1 provides conditions under which (60) holds with $y = Y_n = F_n^{-1}(\Phi(Z))$ and $u(Y_n) = \frac{2K_n Y_n^2}{\sqrt{n}} + \frac{2K_n}{\sqrt{n}}$. For more about this approach see Sakhanenko [49].

4.2. Proofs for the KMT quantile inequality section

4.2.1. Proof of Theorem 1

We shall infer Theorem 1 from the following technical result.

Proposition 3. *Assume there exist a sequence $K_n > 0$, a sequence $0 < \varepsilon_n < 1$ and an integer $n_0 \geq 1$ such that for all $n \geq n_0$ and $0 < z \leq \varepsilon_n \sqrt{n}$ inequalities (10), (11), (12) and (13) hold. Then whenever $n \geq n_0 \vee \left(64K_n^2\right)$ and $|x| \leq \eta_n = \varepsilon_n \wedge (1/(8K_n))$ we have*

$$\Phi(\sqrt{n}x + u) \geq F_n(\sqrt{n}x) \geq F_n(\sqrt{n}x-) \geq \Phi(\sqrt{n}x - u), \tag{63}$$

where

$$u = 2K_n\sqrt{n}x^2 + \frac{2K_n}{\sqrt{n}}. \tag{64}$$

To see how our theorem follows from this proposition, set $Y_n = \sqrt{n}x$ into the (63) and (64). Therefore whenever $n \geq n_0 \vee \left(64K_n^2\right)$ and

$$|Y_n| \leq \eta_n \sqrt{n}, \tag{65}$$

we have

$$\Phi(Y_n + u) \geq F_n(Y_n) \geq F_n(Y_n-) \geq \Phi(Y_n - u),$$

where

$$u = \frac{2K_n Y_n^2}{\sqrt{n}} + \frac{2K_n}{\sqrt{n}}.$$

As pointed out in subsection 4.1, this inequality implies

$$-\frac{2K_n Y_n^2}{\sqrt{n}} - \frac{2K_n}{\sqrt{n}} \leq Y_n - Z \leq \frac{2K_n Y_n^2}{\sqrt{n}} + \frac{2K_n}{\sqrt{n}}.$$

Thus (15) holds. Hence the theorem will be proved as soon as we have established the proposition.

4.2.2. Proof of Proposition 3

The proof will follow from a number of lemmas.

Lemma 1. *For any $x > 0$*

$$\frac{1}{x}\left(1 - \frac{1}{x^2}\right) \leq \frac{1 - \Phi(x)}{\phi(x)} \leq \frac{1}{x}. \tag{66}$$

(This is the classic Mill's ratio. Refer, for instance, to Shorack and Wellner [53].) We can readily infer from (66) and some simple bounds that for some $c > 1$ and all $x \geq 0$,

$$\frac{1}{c\,(1 + x)} \leq \frac{1 - \Phi(x)}{\phi(x)} \leq \frac{c}{1 + x}. \tag{67}$$

Lemma 2. *The function*

$$\Psi_1(x) := \frac{\phi(x)}{1-\Phi(x)} \text{ is increasing} \tag{68}$$

and the function

$$\Psi_2(x) := \frac{\phi(x)}{\Phi(x)} \text{ is decreasing.} \tag{69}$$

Proof. First consider (69). We see that

$$\Psi_2'(x) = \frac{\phi'(x)\Phi(x) - \phi^2(x)}{\Phi^2(x)}.$$

Now

$$\phi'(x)\Phi(x) - \phi^2(x) = (2\pi)^{-1}\exp(-x^2)g(x),$$

where

$$g(x) = -x\exp(x^2/2)\int_{-\infty}^{x}\exp(-t^2/2)dt - 1.$$

When $x \geq 0$, obviously $g(x) < 0$, and when $x < 0$, (66) implies that $g(x) \leq 0$. Thus we have (69). Assertion (68) follows from the fact that $\Psi_1(x) := \phi(-x)/\Phi(-x) = \Psi_2(-x)$. □

Notice that (66) and (68) imply that for $x > 0$, we have

$$\frac{\phi(x)}{1-\Phi(x)} > \max\left\{x, \frac{2}{\sqrt{2\pi}}\right\} \geq \frac{1}{2}\left(x + \frac{2}{\sqrt{2\pi}}\right). \tag{70}$$

The following lemma can be inferred from Lemma (A.8) in Einmahl [18] with the 8 there replaced by an unspecified constant. To keep the presentation self-contained we provide here a direct proof. In any case, we need this lemma with its present constants in order to establish Theorem 1 as it is stated.

Lemma 3. *For all* $0 < A < \infty$, $n \geq 64A^2$ *and* $0 \leq x \leq 1/(8A)$ *we have*

$$\log\left(\frac{\Phi(-\sqrt{n}x + u)}{\Phi(-\sqrt{n}x)}\right) = \log\left(\frac{1-\Phi(\sqrt{n}x - u)}{1-\Phi(\sqrt{n}x)}\right) \geq A(nx^3 + n^{-1/2}) \tag{71}$$

and

$$\log\left(\frac{\Phi(-\sqrt{n}x - u)}{\Phi(-\sqrt{n}x)}\right) = \log\left(\frac{1-\Phi(\sqrt{n}x + u)}{1-\Phi(\sqrt{n}x)}\right) \leq -A(nx^3 + n^{-1/2}), \tag{72}$$

where $u = 2A(\sqrt{n}x^2 + n^{-1/2})$.

Proof. We shall treat (71) first. Notice that by the mean value theorem

$$\log\left(\frac{1-\Phi(\sqrt{n}x - u)}{1-\Phi(\sqrt{n}x)}\right) = \frac{u\phi(\xi)}{1-\Phi(\xi)}, \tag{73}$$

where $\xi \in [\sqrt{n}x - u, \sqrt{n}x]$.

Choose any $A > 0$.

(i) First assume that $0 \le x \le n^{-1/2}$.

Obviously we have $\xi \in [-4A/\sqrt{n}, 1]$. Thus by (68)

$$\frac{u\phi(\xi)}{1-\Phi(\xi)} \ge \frac{u\phi(-4A/\sqrt{n})}{1-\Phi(-4A/\sqrt{n})} = 2A(\sqrt{n}x^2 + n^{-1/2})\Psi_1(-4A/\sqrt{n}),$$

which since $\sqrt{n}x \le 1$ is

$$\ge 2A(nx^3 + n^{-1/2})\Psi_1(-4A/\sqrt{n}). \tag{74}$$

(ii) Now assume that $(8A)^{-1} \ge x > n^{-1/2}$. Since $x^2 n > 1$, we have

$$\sqrt{n}x - u = \sqrt{n}x - 2A(\sqrt{n}x^2 + n^{-1/2}) \ge \sqrt{n}x[1-4Ax] \ge \sqrt{n}x/2 > 0, \tag{75}$$

so we can apply (66) to get

$$\frac{\phi(\sqrt{n}x - u)}{1-\Phi(\sqrt{n}x-u)} \ge \sqrt{n}x - u.$$

This combined with (68), (73) and (75) gives

$$\log\left(\frac{1-\Phi(\sqrt{n}x-u)}{1-\Phi(\sqrt{n}x)}\right) \ge u(\sqrt{n}x - u)$$

$$\ge u\sqrt{n}x/2 = A(nx^3 + x) \ge A(nx^3 + n^{-1/2}). \tag{76}$$

Choose any integer n such that $-4A/\sqrt{n} \ge -1/2$. Notice that for such n we also have

$$\Psi_1(-4A/\sqrt{n}) \ge \Psi_1(-1/2) = .5091 > .5. \tag{77}$$

Hence by inequalities (74), (76) and (77) we see that for all $n \ge 64A^2$ and $0 \le x \le 1/(8A)$ (see (i) and (ii)) inequality (71) holds.

Next consider (72). Choose $n \ge 64A^2$ and $0 \le x \le 1/(8A)$. We need only show that

$$\log\left(\frac{1-\Phi(\sqrt{n}x)}{1-\Phi(\sqrt{n}x+u)}\right) \ge \log\left(\frac{1-\Phi(\sqrt{n}x-u)}{1-\Phi(\sqrt{n}x)}\right). \tag{78}$$

Now

$$\log\left(\frac{1-\Phi(\sqrt{n}x-u)}{1-\Phi(\sqrt{n}x)}\right) = \frac{u\phi(\xi_1)}{1-\Phi(\xi_1)},$$

where $\xi_1 \in [\sqrt{n}x - u, \sqrt{n}x]$ and

$$\log\left(\frac{1-\Phi(\sqrt{n}x)}{1-\Phi(\sqrt{n}x+u)}\right) = \frac{u\phi(\xi_2)}{1-\Phi(\xi_2)},$$

where $\xi_2 \in [\sqrt{n}x, \sqrt{n}x + u]$.

Since $\xi_1 \le \xi_2$, (78) follows from (68). Assertion (72) follows from (71). $\square$

We are now ready to complete the proof of Proposition 3. Suppose $n \geq n_0 \vee \left(64K_n^2\right)$. By assumptions (10), (11), (12), (13) and Lemma 3 we can choose $A = K_n$, and $\eta_n = \varepsilon_n \wedge (1/(8K_n))$ such that for all $0 \leq x \leq \eta_n$ we have

$$F_n(-\sqrt{n}x) \geq \Phi(-\sqrt{n}x)\exp(-K_n(nx^3 + n^{-1/2})), \tag{79}$$

$$F_n(-\sqrt{n}x) \leq \Phi(-\sqrt{n}x)\exp(K_n(nx^3 + n^{-1/2})), \tag{80}$$

$$1 - F_n(\sqrt{n}x) \geq (1 - \Phi(\sqrt{n}x))\exp(-K_n(nx^3 + n^{-1/2})), \tag{81}$$

$$1 - F_n(\sqrt{n}x) \leq (1 - \Phi(\sqrt{n}x))\exp(K_n(nx^3 + n^{-1/2})), \tag{82}$$

and (71) and (72) hold, which imply that for all $0 \leq x \leq \eta_n$

$$\Phi(-\sqrt{n}x + u) \geq F_n(-\sqrt{n}x) \geq F_n(-\sqrt{n}x-) \geq \Phi(-\sqrt{n}x - u) \tag{83}$$

and

$$1 - \Phi(\sqrt{n}x - u) \geq 1 - F_n(\sqrt{n}x-) \geq 1 - F_n(\sqrt{n}x) \geq 1 - \Phi(\sqrt{n}x + u). \tag{84}$$

In other words, for all $|x| \leq \eta_n$, we have

$$\Phi(\sqrt{n}x + u) \geq F_n(\sqrt{n}x) \geq F_n(\sqrt{n}x-) \geq \Phi(\sqrt{n}x - u),$$

where $u = 2K_n\sqrt{n}x^2 + \frac{2K_n}{\sqrt{n}}$. This completes the proof of the proposition. □

4.2.3. Proof of Corollary 1

We know by Theorem 2 that there exist an $0 < L < \infty$, an $0 < \eta < \infty$ and an integer $n_0 \geq 1$ such that for all integers $n \geq n_0 \vee \left(64L^2 n^{1-2/p}\right)$, whenever (21) holds, we have

$$n^{1/p}\left|Y_n - Z\right| \leq 2LY_n^2 + 2L. \tag{85}$$

We require a number of lemmas.

Lemma 4. *For every $A > 0$ there exist $C_1 > 0$ and $\lambda_1 > 0$ such that for all $n \geq 1$ and $0 \leq z \leq An^{2/p}$ we have*

$$P\left\{Y_n^2 > z\right\} \leq C_1 e^{-\lambda_1 z}. \tag{86}$$

Proof. By (23) we have

$$\begin{aligned} P\left\{Y_n^2 > z\right\} &= P\left\{|Y_n| > z^{1/2}\right\} \\ &\leq c\exp\left(-bz\left(1 + a\left(n^{-1/p}\sqrt{z}\right)^{2p/(p+2)}\right)^{-1}\right) \\ &\leq c\exp\left(-bz\left(1 + aA^{p/(p+2)}\right)^{-1}\right). \end{aligned}$$

Set $C_1 = c$ and $\lambda_1 = b\left(1 + aA^{p/(p+2)}\right)^{-1}$. □

Lemma 5. *For every $A > 0$ there exist positive numbers $C_2 > 0$ and $\lambda_2 > 0$ such that for all $n \geq 1$, $k \geq 1$ and $0 \leq z \leq An^{2/p}$ we have*

$$P\left\{|Y_n| > \eta n^{1/p}\right\} \leq C_2 \exp\left(-z\lambda_2\right). \tag{87}$$

Proof. Applying (23), we see that

$$P\left\{|Y_n| > \eta n^{1/p}\right\} \leq c \exp\left(-\frac{b\eta^2 n^{2/p}}{1 + a\eta^{2p/(p+2)}}\right),$$

which for $C_2 = c$, $\lambda_2 = A^{-1}b\eta^2\left(1 + a\eta^{2p/(p+2)}\right)^{-1}$ and any $0 < z \leq An^{2/p}$ and all $n \geq 1$ is

$$\leq C_2 \exp\left(-\lambda_2 z\right).$$

□

Combining Lemmas 4 and 5 with inequality (22) we readily infer the following lemma:

Lemma 6. *For every $A > 0$ there exist positive numbers $C_3 > 0$ and $\lambda_3 > 0$ such that for all $n \geq n_0 \vee \left(64K^2\right)$ and $0 \leq z \leq An^{2/p}$ we have*

$$P\left\{n^{1/p}\left|Y_n - Z\right| > z\right\} \leq C_3 \exp(-\lambda_3 z). \tag{88}$$

We shall need two more lemmas.

Lemma 7. *For every $A > 0$ there exist positive numbers $C_4 > 0$ and $\lambda_4 > 0$ such that for all $n \geq 1$, and $z > An^{2/p}$ we have*

$$P\left\{\left|n^{1/p} Y_n\right| > z\right\} \leq C_4 \exp(-\lambda_4 z^{4/(p+2)}). \tag{89}$$

Proof. The proof is an easy consequence of inequality (23), which gives

$$P\left\{|Y_n| > z/n^{1/p}\right\} \leq c \exp\left(-\frac{bz^2/n^{2/p}}{1 + a\left(n^{-2/p}z\right)^{2p/(p+2)}}\right)$$

$$\leq c \exp\left(-\frac{bz^2/n^{2/p}}{2}\right) + c \exp\left(-\frac{bz^{2-2p/(p+2)}}{2an^{-4/(p+2)+2/p}}\right)$$

which since $z \geq An^{2/p}$ is

$$\leq c \exp\left(-\frac{bzA}{2}\right) + c \exp\left(-\frac{bz^{4/(p+2)}\left(n^{2/p}\right)^{(p-2)/(p+2)}}{2a}\right),$$

which, in turn, since $p \geq 2$ is

$$\leq c \exp\left(-\frac{bzA}{2}\right) + c \exp\left(-\frac{bz^{4/(p+2)}}{2a}\right).$$

□

Lemma 8. *For every $A > 0$ there exist positive numbers $C_5 > 0$ and $\lambda_5 > 0$ such that for all $n \geq 1$, and $z > An^{2/p}$ we have*

$$P\left\{\left|n^{1/p}Z\right| > z\right\} \leq C_5 \exp(-\lambda_5 z). \tag{90}$$

Proof. Inequality (90) is readily inferred from the elementary bound

$$P\left\{\left|n^{1/p}Z\right| > z\right\} \leq 2\exp\left(-z^2/\left(2n^{2/p}\right)\right), \text{ for all } z \geq 0.$$

□

The proof of inequality (24) for $n \geq n_0$ is now completed by using some routine bounds on the probability in (24) and then applying Lemmas 6, 7 and 8. For the case $1 \leq n < n_0$, (should it be that $n_0 > 1$), we establish that (24) holds uniformly in $1 \leq n < n_0$ by using the elementary inequality

$$P\left\{n^{1/p}\left|Y_n - Z\right| > z\right\} \leq P\left\{n_0^{1/p}\left|Y_n\right| > z/2\right\} + P\left\{n_0^{1/p}\left|Z\right| > z/2\right\}.$$

□

Remark 7. Actually the proof shows that for suitable constants $C > 0$ and $\lambda > 0$,

$$P\left\{n^{1/p}\left|Y_n - Z\right| > z\right\}$$
$$\leq C\left(\exp\left(-\lambda z\right) \vee \exp\left(-\lambda z^{4/(p+2)}\left(n^{2/p}\right)^{(p-2)/(p+2)}\right)\right).$$

4.2.4. Proof of Corollary 2

Assumptions (25) and (26) allow us to apply the results in Grama and Haeusler [22] to get the following large-deviation result: For x in the range $1 \leq x \leq \alpha_+ n^{1/4}$ (for $\alpha_+ > 0$ sufficiently small), one has

$$\frac{P\{Y_n > x\}}{1 - \Phi(x)} = \exp\left(O\left(\frac{x^3}{\sqrt{n}}\right)\right)\left\{1 + O\left(\frac{(M+L)x\log n}{\sqrt{n}}\right)\right\}.$$

This implies that for some constant $D_1 > 0$, all large enough n and $1 \leq x \leq \alpha_+ n^{1/4}$

$$\exp\left(-D_1 \log n\left(\frac{x^3+1}{\sqrt{n}}\right)\right) \leq \frac{P\{Y_n > x\}}{1 - \Phi(x)} \leq \exp\left(D_1 \log n\left(\frac{x^3+1}{\sqrt{n}}\right)\right).$$

Furthermore assumptions (25), (26) and (27) permit us to apply the corollary in Bolthausen [5] to infer that for some constant $D_2 > 0$ for all $n \geq 2$

$$\sup_{x\in\mathbb{R}}\left|P\{Y_n > x\} - (1 - \Phi(x))\right| \leq \frac{D_2 \log n}{\sqrt{n}}.$$

This implies that for some $D_3 > 0$ for all $0 \le x \le 1$

$$\exp\left(\frac{-D_3 \log n}{\sqrt{n}}\right) \le \frac{P\{Y_n > x\}}{1-\Phi(x)} \le \exp\left(\frac{D_3 \log n}{\sqrt{n}}\right).$$

Thus for some $D_+ > 0$, all large enough n and $0 \le x \le \alpha_+ n^{1/4}$

$$\exp\left(-D_+ \log n \left(\frac{x^3+1}{\sqrt{n}}\right)\right) \le \frac{P\{Y_n > x\}}{1-\Phi(x)} \le \exp\left(D_+ \log n \left(\frac{x^3+1}{\sqrt{n}}\right)\right).$$

Similarly we get for some $\alpha_- > 0$, $D_- > 0$, all large enough n and $0 \le x \le \alpha_- n^{1/4}$

$$\exp\left(-D_- \log n \left(\frac{x^3+1}{\sqrt{n}}\right)\right) \le \frac{P\{Y_n < -x\}}{\Phi(-x)} \le \exp\left(D_- \log n \left(\frac{x^3+1}{\sqrt{n}}\right)\right).$$

Thus (10), (11), (12) and (13) are satisfied with $K_n = D \log n$ for some $D > 0$ and with $\varepsilon_n = \alpha/\sqrt[4]{n}$ for some $\alpha > 0$, and for all $n \ge n_0$ for some integer $n_0 \ge 2$. Applying Theorem 1, whenever $n \ge n_0 \vee (64D^2 (\log n)^2)$, we get (29), namely,

$$|Y_n - Z| \le 2D \log n \left(Y_n^2 + 1\right)/\sqrt{n},$$

whenever $|Y_n| \le \eta_n \sqrt{n} = \alpha \sqrt[4]{n}$ with $\eta_n = \alpha/\sqrt[4]{n}$. Furthermore, note that $n \ge n_0 \vee (64D^2 (\log n)^2)$ for all $n \ge n_1$ for some $n_1 \ge 2$.

Next consider (30). Notice that for $n \ge 2$

$$\sqrt[4]{n}\,|Y_n - Z| \le \sqrt{n}\,|Y_n - Z| / \log n,$$

which for $|Y_n| \le \alpha \sqrt[4]{n}$ and $n \ge n_1$ is

$$\le 2D\left(Y_n^2 + 1\right).$$

Also

$$P\left\{Y_n^2 > z\right\} = P\left\{|Y_n| > \sqrt{z}\right\},$$

which by Azuma's inequality is

$$\le 2\exp\left(-z/\left(2L^2\right)\right).$$

Applying Azuma's inequality again we have

$$P\left\{\sqrt{n}\,|Y_n| > \sqrt{n}\alpha\sqrt[4]{n}\right\} \le 2\exp\left(-\alpha^2\sqrt{n}/\left(2L^2\right)\right), \tag{91}$$

which for $0 \le z \le \sqrt{n}$ is

$$\le C\exp\left(-\alpha^2 z/\left(2L^2\right)\right).$$

Once more using Azuma's inequality we get for $z > \sqrt{n}$

$$P\left\{\sqrt[4]{n}\,|Y_n| > z/2\right\} \le 2\exp\left(-z^2/\left(8L^2\sqrt{n}\right)\right) \le 2\exp\left(-z/\left(8L^2\right)\right)$$

and an elementary bound gives for $z > \sqrt{n}$

$$P\left\{\sqrt[4]{n}\,|Z| > z/2\right\} \le 2\exp\left(-z^2/\left(4\sqrt{n}\right)\right) \le 2\exp\left(-z/4\right).$$

It is now easy to conclude from these inequalities that (30) holds for appropriate $C > 0$ and $\lambda > 0$. $\square$

4.3. Proofs for the refined quantile inequalities section

4.3.1. Proof of Theorem 4

The following is a detailed proof of Theorem 4. It is a modification of the proof for the classical case, which was sketched in Komlós, Major, and Tusnády [31].

Recall the definition of Y_n as in (4). Without loss of generality, we assume that $0 \le Y_n \le \varepsilon\sqrt{n}$, because the derivation for $-\varepsilon\sqrt{n} \le Y_n \le 0$ is similar. Equation (39) is equivalent to

$$-C_1 \frac{1}{n}\left(1+|Y_n|^3\right) \le Y_n - Z \le C_1 \frac{1}{n}\left(1+|Y_n|^3\right),$$

i.e.

$$\Phi\left(Y_n - C_1 \frac{1}{n}\left(1+|Y_n|^3\right)\right) \le \Phi(Z) \le \Phi\left(Y_n + C_1 \frac{1}{n}\left(1+|Y_n|^3\right)\right).$$

Define $F_n(x-) = P(Y_n < x)$. From the definition of Y_n in (4) we have $F_n(Y_n-) \le \Phi(Z) \le F_n(Y_n)$. Thus we need only to show

$$\begin{aligned} &\Phi\left(Y_n - C_1 \frac{1}{n}\left(1+|Y_n|^3\right)\right) \\ &\le F_n(Y_n-) \le F_n(Y_n) \le \Phi\left(Y_n + C_1 \frac{1}{n}\left(1+|Y_n|^3\right)\right), \end{aligned} \tag{92}$$

i.e.

$$\begin{aligned} &\log\left(\frac{1-\Phi\left(x - C_1\frac{1}{n}\left(1+x^3\right)\right)}{1-\Phi(x)}\right) \\ &\ge \log\frac{1-F_n(x-)}{1-\Phi(x)} \ge \log\frac{1-F_n(x)}{1-\Phi(x)} \\ &\ge \log\left(\frac{1-\Phi\left(x + C_1\frac{1}{n}\left(1+x^3\right)\right)}{1-\Phi(x)}\right), \end{aligned}$$

whenever $0 \le x \le \varepsilon\sqrt{n}$.

Lemma 9 shows that Equation (92) holds when $n \le n_0$ for any fixed integer $n_0 \ge 1$. It is then enough to consider the n sufficiently large case. From the assumptions of Theorem 4, we know that for $0 \le x \le \varepsilon\sqrt{n}$,

$$\max\left\{\left|\log\frac{1-F_n(x)}{1-\Phi(x)}\right|, \left|\log\frac{1-F_n(x-)}{1-\Phi(x)}\right|\right\} \le C\left(n^{-1}x^4 + n^{-1}\right)$$

for some $C > 0$. Thus it suffices to show there is a $C_1 > 0$ and a small enough $\varepsilon_1 > 0$ such that uniformly in $0 \le x \le \varepsilon_1\sqrt{n}$,

$$\log\left(\frac{1-\Phi\left(x - C_1\frac{1}{n}\left(1+x^3\right)\right)}{1-\Phi(x)}\right) \ge C\left(n^{-1}x^4 + n^{-1}\right) \tag{93}$$

and

$$-C\left(n^{-1}x^4+n^{-1}\right) \geq \log\left(\frac{1-\Phi\left(x+C_1\frac{1}{n}\left(1+x^3\right)\right)}{1-\Phi\left(x\right)}\right).$$

We shall only show the first part of the inequality above due to the symmetry of the equation. (See the proof of inequality (72) in Lemma 3.)

We shall first verify that the first part of the equation above is satisfied under the condition $x - C_1\frac{1}{2n}\left(1+x^3\right) \leq 0$ and $0 \leq x \leq \varepsilon_1\sqrt{n}$ (we will see later the value of C_1 can be specified as $18\sqrt{2\pi}C$). It implies $x \leq C_1/n \leq 1$ for n sufficiently large under the assumption that $C_1\varepsilon_1^2 \leq 1$, which holds by choosing $\varepsilon_1 > 0$ sufficiently small. Then for $0 \leq x \leq C_1/n \leq 1$ and n sufficiently large, we have

$$\begin{aligned}\log\left(\frac{1-\Phi\left(x-C_1\frac{1}{n}\left(1+x^3\right)\right)}{1-\Phi\left(x\right)}\right) &\geq \log\left(\frac{1-\Phi\left(x-C_1\frac{1}{n}\left(1+x^3\right)\right)}{1-\Phi\left(0\right)}\right)\\ &= \log\left(1+\left[1-2\Phi\left(x-C_1\frac{1}{n}\left(1+x^3\right)\right)\right]\right)\\ &\geq \frac{1}{2}-\Phi\left(x-C_1\frac{1}{n}\left(1+x^3\right)\right),\end{aligned}$$

where the last inequality follows from the fact $\log\left(1+y\right) \geq y/2$ for $0 \leq y \leq 1$. Write

$$\frac{1}{2}-\Phi\left(x-C_1\frac{1}{n}\left(1+x^3\right)\right) = \Phi\left(C_1\frac{1}{n}\left(1+x^3\right)-x\right)-\Phi\left(0\right).$$

Since $0 \leq C_1\frac{1}{n}\left(1+x^3\right) \leq 2$ and $\phi\left(u\right) \geq \frac{1}{9\sqrt{2\pi}}$ for $0 \leq u \leq 2$, and keeping in mind that we are assuming that $x - C_1\frac{1}{2n}\left(1+x^3\right) \leq 0$, the mean value theorem implies

$$\begin{aligned}\Phi\left(C_1\frac{1}{n}\left(1+x^3\right)-x\right)-\Phi\left(0\right) &\geq \frac{1}{9\sqrt{2\pi}}\left(C_1\frac{1}{n}\left(1+x^3\right)-x\right)\\ &\geq \frac{1}{9\sqrt{2\pi}}\cdot\frac{C_1}{2n}\left(1+x^3\right) \geq \frac{C_1}{18\sqrt{2\pi}}\left(n^{-1}x^4+n^{-1}\right),\end{aligned}$$

which is more than $C\left(n^{-1}x^4+n^{-1}\right)$ whenever $C_1 \geq 18\sqrt{2\pi}C$. Thus equation (93) is established in the case of $x - C_1\frac{1}{2n}\left(1+x^3\right) \leq 0$.

Now we consider the case $x - C_1\frac{1}{2n}\left(1+x^3\right) \geq 0$. The mean value theorem tells us there is a number ξ between x and $x - \frac{C_1}{4n}\left(1+x^3\right)$ such that

$$\begin{aligned}&\log\left(\frac{1-\Phi\left(x-C_1\frac{1}{n}\left(1+x^3\right)\right)}{1-\Phi\left(x\right)}\right)\\ &\geq \log\left(\frac{1-\Phi\left(x-\frac{C_1}{4n}\left(1+x^3\right)\right)}{1-\Phi\left(x\right)}\right)\\ &= \frac{C_1}{4}\frac{1}{n}\left(1+x^3\right)\frac{\phi\left(\xi\right)}{1-\Phi\left(\xi\right)}.\end{aligned}$$

From (70) and (68) we have

$$\log\left(\frac{1-\Phi\left(x-\frac{C_1}{4n}\left(1+x^3\right)\right)}{1-\Phi\left(x\right)}\right)$$
$$\geq \frac{C_1}{4n}\left(1+x^3\right)\cdot\frac{1}{2}\left(x-\frac{C_1}{4n}\left(1+x^3\right)+\frac{2}{\sqrt{2\pi}}\right)$$
$$\geq \frac{C_1}{4}\frac{1}{n}\left(1+x^3\right)\cdot\frac{1}{2}\left(\frac{x}{2}+\frac{2}{\sqrt{2\pi}}\right)\geq\frac{C}{n}x^4+\frac{C}{n},$$

whenever $C_1 \geq 16C$.

Putting everything together, we establish that (93) holds for all large enough $n \geq 1$. To complete the proof recall that we assume (37), which allows us to apply Lemma 9 in the appendix to conclude that (93) holds for all $1 \leq n \leq n_0$ for any fixed $n_0 \geq 1$. This finishes the proof of the theorem. □

4.3.2. Proof of Corollary 4

Set $Y_n(z) = H_n(\Phi(z))$. Let us rewrite (39) as

$$|Y_n(z)-z| \leq \frac{C_1}{n}+\frac{C_1}{n}|Y_n(z)|^3, \tag{94}$$

whenever $|Y_n(z)| \leq \varepsilon_1\sqrt{n}$. Obviously inequality (94) still holds, whenever for $0<\varepsilon_2\leq\varepsilon_1$, $|Y_n(z)|\leq\varepsilon_2\sqrt{n}$. Let us choose ε_2 small enough such that $C_1\varepsilon_2^2 < 1/2$. When $|Y_n(z)|\leq\varepsilon_2\sqrt{n}$, we have from (94)

$$|Y_n(z)-z| \leq \frac{C_1}{n}+\frac{1}{2}|Y_n(z)|,$$

which implies by the triangle inequality

$$|Y_n(z)|-|z| \leq \frac{C_1}{n}+\frac{1}{2}|Y_n(z)|,$$

i.e.,

$$|Y_n(z)| \leq \frac{2C_1}{n}+2|z|, \tag{95}$$

so we have for some $C_2>0$,

$$|Y_n(z)-z| \leq \frac{C_1}{n}+\frac{C_1}{n}\left(\frac{2C_1}{n}+2|z|\right)^3 \leq \frac{C_2}{n}\left(1+|z|^3\right).$$

Suppose $Y_n(z_n) = \varepsilon_n(z_n)\sqrt{n} > 0$ with $0<\varepsilon_n(z_n)\leq\varepsilon_2$. We know that $z_n \geq 0$ from the definition of the quantile coupling $Y_n(z) = H_n(\Phi(z))$, and from (95) we have

$$z_n \geq \frac{\varepsilon_n(z_n)}{2}\sqrt{n}-\frac{C_1}{n}.$$

We also see that since $Y_n(z)$ is an increasing function of z, we have $Y_n(z) \le \varepsilon_n(z_n)\sqrt{n} \le \varepsilon_2\sqrt{n}$, whenever $z \le \frac{\varepsilon_n(z_n)}{2}\sqrt{n} - \frac{C_1}{n}$. Similarly we may show $Y_n(z) \ge -\varepsilon_n(z_n)\sqrt{n} \ge -\varepsilon_2\sqrt{n}$, whenever $z \ge -\frac{\varepsilon_n(z_n)}{2}\sqrt{n} + \frac{C_1}{n}$. Thus

$$|Y_n(z)| \le \varepsilon_2\sqrt{n}, \text{ whenever } |z| \le \frac{\varepsilon_n(z_n)}{2}\sqrt{n} - \frac{C_1}{n}. \tag{96}$$

Since $Y_n \to_d N(0,1)$ we know that for all $n \ge n_0$ for some $n_0 \ge 1$ we can choose $\varepsilon_2 \ge \varepsilon_n(z_n) \ge \varepsilon_3 = \varepsilon_2/2$. Hence

$$|Y_n(z)| \le \varepsilon_2\sqrt{n}, \text{ whenever } |z| \le \frac{\varepsilon_3}{2}\sqrt{n} - \frac{C_1}{n}. \tag{97}$$

Let $\varepsilon_4 = \varepsilon_3/4$. We have $\varepsilon_4\sqrt{n} < \frac{\varepsilon_3}{2}\sqrt{n} - \frac{C_1}{n}$ for $n \ge n_1 = n_0 \vee \left(\frac{4C_1}{\varepsilon_3}\right)^{2/3}$. Thus

$$\left\{|z| \le \varepsilon_4\sqrt{n}\right\} \subset \left\{|z| \le \frac{\varepsilon_3}{2}\sqrt{n} - \frac{C_1}{n}\right\} \subset \left\{|Y_n(z)| \le \varepsilon_2\sqrt{n}\right\},$$

so we have

$$|Y_n(z) - z| \le \frac{C_2}{n}\left(1 + |z|^3\right), \text{ whenever } |z| \le \varepsilon_4\sqrt{n} \text{ and } n \ge n_1.$$

When $|z| \le \varepsilon_4\sqrt{n}$ and $n \le n_1$, we can infer from (95) that

$$|Y_n(z)| \le \frac{2C_1}{n} + 2|z| \le \frac{2C_1}{n} + 2\varepsilon_4\sqrt{n} \le C_3/n$$

with $C_3 = 2C + 2\varepsilon_4(n_1)^{3/2}$. Thus we get

$$|Y_n(z) - z| \le \frac{C_3}{n}\left(1 + |z|^3\right), \text{ whenever } |z| \le \varepsilon_4\sqrt{n}.$$

Setting $C = C_1 \vee C_2$ and $\varepsilon = \varepsilon_4$ completes the proof. □

4.3.3. Proof of Theorem 5

The proof of Theorem 5 follows the same lines as that of Theorem 4. Repeating the first part of the proof of Theorem 4 with $\frac{1}{n} + \frac{|Y_n|^3}{n}$ replaced by $\frac{1}{\sqrt{n}} + \frac{|Y_n|^3}{n}$ and $\frac{1}{n} + \frac{|x|^3}{n}$ by $\frac{1}{\sqrt{n}} + \frac{|x|^3}{n}$ and using the fact that by the assumptions of Theorem 5 for some $C > 0$, uniformly in $n \ge 1$ and for all $x \in [0, \varepsilon\sqrt{n}] \cap \mathbb{C}_n$

$$\max\left|\log\frac{1 - F_n(x-)}{1 - \Phi(x)}, \log\frac{1 - F_n(x)}{1 - \Phi(x)}\right| \le C\left(n^{-1}x^4 + n^{-1/2}\right),$$

we see that to complete the proof it is enough to show there is $C_1 > 0$ such uniformly in $x \in [0, \varepsilon_1\sqrt{n}]$, for a small enough $\varepsilon_1 > 0$ all large enough n

$$\log\left(\frac{1 - \Phi\left(x - C_1\left(\frac{1}{\sqrt{n}} + \frac{x^3}{n}\right)\right)}{1 - \Phi(x)}\right) \ge C\left(n^{-1}x^4 + n^{-1/2}\right)$$

and

$$-C\left(n^{-1}x^4+n^{-1/2}\right)\geq \log\left(\frac{1-\Phi\left(x+C_1\left(\frac{1}{\sqrt{n}}+\frac{x^3}{n}\right)\right)}{1-\Phi(x)}\right).$$

The remainder of the proof is then nearly identical to that of Theorem 4. Replace everywhere in this part of the proof of Theorem 4, $C_1\left(\frac{1}{n}+\frac{x^3}{n}\right)$ by $C_1\left(\frac{1}{\sqrt{n}}+\frac{x^3}{n}\right)$ and note that now when $x-\frac{C_1}{2}\left(\frac{1}{\sqrt{n}}+\frac{x^3}{n}\right)\leq 0$, $0\leq x\leq \varepsilon_1\sqrt{n}$ and $C_1\varepsilon_1^2\leq 1$, then $0\leq x\leq C_1/\sqrt{n}\leq 1$ for n sufficiently large. The rest of the details are the same. □

Appendix A

This result is an extension of Lemma 1.2.1 of Mason [37], where Binomial distributions were considered.

Lemma 9. *Under condition (37), for any fixed $n_0\geq 1$ there exist $0<\varepsilon<\infty$ and $0<C<\infty$ such that for all $x\in\left[-\varepsilon\sqrt{n_0},\varepsilon\sqrt{n_0}\right]$ and $1\leq n\leq n_0$ we have*

$$\Phi\left(x-C\frac{1}{n}\left(1+|x|^3\right)\right)\leq F_n(x-)\leq F_n(x)\leq \Phi\left(x+C\frac{1}{n}\left(1+|x|^3\right)\right), \quad (98)$$

$$\exp\left(-C\left(x^4+1\right)/n\right)\leq \frac{1-F_n(x)}{1-\Phi(x)}\leq \exp\left(C\left(x^4+1\right)/n\right) \quad (99)$$

and

$$\exp\left(-C\left(x^4+1\right)/n\right)\leq \frac{F_n(x-)}{\Phi(x-)}\leq \exp\left(C\left(x^4+1\right)/n\right). \quad (100)$$

Proof. By (37) for any $1\leq n\leq n_0$ we can choose an $\varepsilon_n>0$ such that

$$0<F_n(-\varepsilon_n-)\leq F_n(\varepsilon_n)<1.$$

Define $\varepsilon=\frac{1}{\sqrt{n_0}}\left(\frac{1}{2}\wedge\min_{n\leq n_0}\varepsilon_n\right)$. Then for all $x\in\left[-\varepsilon\sqrt{n_0},\varepsilon\sqrt{n_0}\right]$ and $1\leq n\leq n_0$ we have

$$\begin{aligned}0&<\min_{n\leq n_0}F_n(-\varepsilon_n-)\leq F_n(-\varepsilon_n-)\leq F_n(x-)\\&\leq F_n(x)\leq F_n(\varepsilon_n)\leq \max_{n\leq n_0}F_n(\varepsilon_n)<1.\end{aligned}$$

Choose C sufficiently large such that

$$\Phi\left(\frac{1}{2}-\frac{C}{n_0}\right)<\min_{n\leq n_0}F_n(-\varepsilon_n-)\leq \max_{n\leq n_0}F_n(\varepsilon_n)<\Phi\left(-\frac{1}{2}+\frac{C}{n_0}\right),$$

$$\exp(-C/n_0)<\frac{1-\max_{n\leq n_0}F_n(\varepsilon_n)}{1-\Phi\left(-\frac{1}{2}\right)}\leq \frac{1-\min_{n\leq n_0}F_n(-\varepsilon_n-)}{1-\Phi\left(\frac{1}{2}\right)}<\exp(C/n_0)$$

and

$$\exp\left(-C/n_0\right) < \frac{\min_{n\le n_0} F_n\left(-\varepsilon_n-\right)}{\Phi\left(\frac{1}{2}\right)} \le \frac{\max_{n\le n_0} F_n\left(\varepsilon_n\right)}{\Phi\left(-\frac{1}{2}\right)} < \exp\left(C/n_0\right).$$

Since for any $x \in \left[-\varepsilon\sqrt{n_0}, \varepsilon\sqrt{n_0}\right]$ and $1 \le n \le n_0$,

$$x - C\frac{1}{n}\left(1+|x|^3\right) \le \frac{1}{2} - \frac{C}{n_0}, \text{ and } x + C\frac{1}{n}\left(1+|x|^3\right) \ge -\frac{1}{2} + \frac{C}{n_0},$$

we have established (98). Similarly from the fact

$$-C\left(x^4+1\right)/n \le -C/n_0, \text{ i.e., } C\left(x^4+1\right)/n > C/n_0,$$

we get (99) and (100). □

Remark 8. It is easy to see from the proof that Lemma 9 can be generalized with $\Phi\big(x + C\frac{1}{n}(1+|x|^3)\big)$ in (98) replaced by $\Phi\left(x^{a_1} \pm C\frac{1}{n^{a_2}}\left(1+|x|^{a_3}\right)\right)$ and $\exp\left(C\left(x^4+1\right)/n\right)$ in (99) and (100) replaced by $\exp\left(\pm C\left(|x|^{a_4}+1\right)/n^{a_5}\right)$, where constants $a_i > 0$, $1 \le i \le 5$, which implies the requirement "$n \ge n_0$" in most of statements of this paper can be replaced by "$n \ge 1$" as long as (37) holds.

Appendix B

In this subsection we prove the equivalence of (45) and (46). We shall assume that X satisfies $EX = 0$, $0 < VarX = 1 < \infty$, and for some $\gamma \ge 0$ and $K \ge 1$,

$$E|X|^m \le (m!)^{1+\gamma} K^{m-2}, \text{ for } m \ge 3. \tag{101}$$

Notice that (101) implies that for $t > 0$,

$$\frac{t^m}{m!} E|X|^{m/(1+\gamma)} \le \frac{t^m}{m!}\left(E|X|^m\right)^{1/(1+\gamma)} \le t^m \left(K^{1/(1+\gamma)}\right)^{m-2}.$$

Thus there exists a $0 < \delta < 1$ such that

$$E\exp\left(\delta|X|^{1/(1+\gamma)}\right) =: C(\delta) < \infty. \tag{102}$$

Next we shall prove that (102) holding for some $0 < \delta < 1$ implies (101) for some $K > 0$. (The argument is similar to the proof of Lemma 3 in Amosova [1].) Towards showing this, note that for any $m \ge 2$

$$E|X|^m \le \left(E|X|^{\lceil m(1+\gamma)\rceil/(1+\gamma)}\right)^{m(1+\gamma)/\lceil m(1+\gamma)\rceil}, \tag{103}$$

which since $E|X|^{\lceil m(1+\gamma)\rceil/(1+\gamma)} \ge 1$ is

$$\le E|X|^{\lceil m(1+\gamma)\rceil/(1+\gamma)}, \tag{104}$$

which, in turn, with $k = \lceil m(1+\gamma) \rceil$ is

$$\leq k!\delta^{-\lceil m(1+\gamma)\rceil} C(\delta) \leq k!\delta^{-m\gamma-m-1} C(\delta). \tag{105}$$

Now by Stirling's formula

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e.$$

Thus

$$k! \leq \sqrt{2\pi k}\left(\frac{k}{e}\right)^k e.$$

Now for $m \geq 3$, $3 \leq m(1+\gamma) \leq k \leq m(1+\gamma)+1$. Hence

$$\sqrt{2\pi k}\left(\frac{k}{e}\right)^k e \leq \sqrt{2\pi m(1+\gamma)+1}\left(\frac{m(1+\gamma)+1}{e}\right)^{m(1+\gamma)+1} e,$$

which is

$$\leq \sqrt{4\pi m(1+\gamma)}\left(\frac{2m(1+\gamma)}{e}\right)^{m(1+\gamma)} e.$$

Here we used the inequality $(x+1)^{x+1} \leq (2x)^x e$ for $x \geq 1$. This last expression is

$$\leq \sqrt{2\pi m}\left(\frac{m}{e}\right)^{m(1+\gamma)} e\sqrt{2(1+\gamma)}\,(2(1+\gamma))^{m(1+\gamma)}$$

$$\leq \left(\sqrt{2\pi m}\left(\frac{m}{e}\right)^m\right)^{(1+\gamma)} e\sqrt{2(1+\gamma)}\,(2(1+\gamma))^{m(1+\gamma)}.$$

Now

$$\sqrt{2\pi m}\left(\frac{m}{e}\right)^m \leq m!.$$

Hence by keeping (103), (104) and (105) in mind, we get

$$E|X|^m \leq$$

$$k!\delta^{-m\gamma-m-1} C(\delta) \leq (m!)^{(1+\gamma)}\left(\frac{e}{\delta}\right)\sqrt{2(1+\gamma)}\,(2(1+\gamma)/\delta)^{m(1+\gamma)} C(\delta).$$

Clearly for a sufficiently large K inequality (101) holds for all $m \geq 3$.

Now assume that (102) holds for a $0 < \delta < 1$, then for all $x \geq 0$

$$P\{|X| \geq x\} = P\left\{\exp\left(\delta|X|^{1/(1+\gamma)}\right) \geq \exp\left(\delta|x|^{1/(1+\gamma)}\right)\right\}$$

$$\leq C(\delta)\exp\left(-\delta|x|^{1/(1+\gamma)}\right).$$

Going the other way, suppose that for some $\gamma \geq 0$, $C > 0$ and $d > 0$,

$$P\{|X| \geq x\} \leq C\exp\left(-d|x|^{1/(1+\gamma)}\right).$$

Then for any $0 < \delta < 1$

$$\begin{aligned}
E\exp\left(\delta\left|X\right|^{1/(1+\gamma)}\right) &= \int_0^\infty P\left\{\exp\left(\delta\left|X\right|^{1/(1+\gamma)}\right) \geq x\right\}dx \\
&= \int_0^\infty P\left\{\left|X\right|^{1/(1+\gamma)} \geq \delta^{-1}\log x\right\}dx \\
&\leq e + \int_e^\infty P\left\{\left|X\right|^{1/(1+\gamma)} \geq \delta^{-1}\log x\right\}dx \\
&= e + \int_e^\infty P\left\{\left|X\right| \geq \left(\delta^{-1}\log x\right)^{1+\gamma}\right\}dx \\
&\leq e + C\int_e^\infty \exp\left(-d\delta^{-1}\log x\right)dx,
\end{aligned}$$

which is finite for a small enough $0 < \delta < 1$. From these considerations we can readily establish the equivalence of (45) and (46).

Appendix C

Wolf [56] extended results of Saulis [50] under a more general moment condition. In the following propositions $X, X_1, X_2, \ldots, X_n$ are i.i.d. random variables with $EX = 0$, $EX^2 = 1$. As above, we use the notation $Y_n =_d \sum_{i=1}^n X_i/\sqrt{n}$ and $Y_n = H_n(\Phi(Z))$, where H_n is the inverse distribution of F_n (the distribution of Y_n), Z is a standard normal random variable and Φ is its distribution function.

Let g be a continuous increasing function on $[0,\infty)$ such that $g(x)x^{-1}$ strictly decreases on $(0,\infty)$. Also assume that there exist $0 < \alpha < 1$ and a constant $C(g) > 0$ depending on g such that for all $x \in (0,\infty)$ and some increasing function $\rho(x)$ on $(0,\infty)$ satisfying $\lim_{x\to\infty}\rho(x) = \infty$, the following inequality holds

$$\rho(x)\ln x \leq g(x) \leq Cx^\alpha.$$

Further assume that

$$E\exp\{g(|X|)\} < \infty. \tag{106}$$

Choose any $k > 1$ and let $\Lambda(n)$ denote the solution to the equation

$$kx^2 = g\left(\sqrt{n}x\right). \tag{107}$$

For example, when $g(x) = dx^\alpha$ for some $\alpha \in (0,1)$ and $d > 0$

$$\Lambda(n) = \left(\frac{d}{k}\right)^{1/(2-\alpha)} n^{\frac{\alpha}{4-2\alpha}}. \tag{108}$$

An argument based on Theorem 6.3 (Wolf [56]) in Saulis and Statulevičius [51] and following the lines of the proofs in subsection 4.3 we get the following refined quantile inequality.

Proposition 4. *Let $X, X_1, X_2, \ldots, X_n$ be i.i.d. random variables for which $EX = 0$, $EX^2 = 1$ and $\limsup_{|t|\to\infty} |E\exp(itX)| < 1$. Assume that the moment condition (106) holds. Then there exists a positive constant ε such that*

$$P\{Y_n < -x\} = \Phi(-x)\exp\left(O\left(n^{-1/2}|x|^3 + n^{-1/2}\right)\right),$$
$$1 - P\{Y_n < -x\} = \overline{\Phi}(x)\exp\left(O\left(n^{-1/2}|x|^3 + n^{-1/2}\right)\right)$$

in the interval $0 \le x \le \varepsilon\Lambda(n)$. Moreover, there exist $C > 0$, $\eta > 0$ and an integer $n_0 \ge 1$ such that for every $n \ge n_0$, whenever

$$|Y_n| \le \eta\Lambda(n),$$

we have

$$|Y_n - Z| \le \frac{C}{\sqrt{n}} + \frac{C}{\sqrt{n}}|Y_n|^2. \tag{109}$$

Similarly, an argument based on Corollary 6.1 (Wolf [56]) in Saulis and Statulevičius [51] gives the following further refined quantile inequality.

Proposition 5. *Let $X, X_1, X_2, \ldots, X_n$ be i.i.d. random variables for which $EX = 0$, $EX^2 = 1$ and $\limsup_{|t|\to\infty} |E\exp(itX)| < 1$. Assume that the moment condition (106) holds, and additionally $EX^3 = 0$. Then there exists a positive constant ε such that*

$$P\{Y_n < -x\} = \Phi(-x)\exp\left(O\left(n^{-1}x^4 + n^{-1}\right)\right),$$
$$1 - P\{Y_n < x\} = \overline{\Phi}(x)\exp\left(O\left(n^{-1}x^4 + n^{-1}\right)\right),$$

in the interval $0 \le x \le \varepsilon\Lambda(n)$. Moreover, there exist C, $\eta > 0$ and an integer $n_0 \ge 1$ such that for every $n \ge n_0$, whenever

$$|Y_n| \le \eta\Lambda(n),$$

we have

$$|Y_n - Z| \le \frac{C}{n} + \frac{C}{n}|Y_n|^3. \tag{110}$$

Notice that specializing to $g(x) = dx^\alpha$ with $\alpha = 4/(p+2)$ and $d > 0$, we get

$$\Lambda(n) = \left(\frac{d}{k}\right)^{(p+2)/(2p)} n^{1/p} =: d_0 n^{1/p}. \tag{111}$$

Now keeping the discussion in Appendix B in mind, we see that when (106) holds with $g(x) = dx^{4/(p+2)}$ that the assumptions of Theorem 3.1 in Saulis and Statulevičius [51] are fulfilled. Therefore we can conclude that Y_n satisfies the Bernstein type inequality (23). Thus under the assumptions of Proposition 4, using the quantile inequality (109), we can readily modify the proof of Corollary 1 to conclude that for positive constants C and λ, for all $z \ge 0$ and $n \ge 1$,

$$P\left\{n^{1/2}|Y_n - Z| > z\right\} \le C\exp\left(-\lambda z^{4/(p+2)}\right). \tag{112}$$

Acknowledgements

Subsection 2.1.1 was added in response to a suggestion by the referee. Also the authors are grateful to Leonard Santana of North-West University, South Africa for assistance in formatting this paper.

References

[1] Amosova, N. N. (1999). The necessity of the Linnik condition in a theorem on probabilities of large deviations. *Journal of Mathematical Sciences* **93** 255–258. MR1449842

[2] Arak, T. V. and Zaitsev, A. Yu. (1988). *Uniform Limit Theorems for Sums of Independent Random Variables*. Translation of *Trudy. Mat. Inst. Steklov* **174** 1986. *Proc. Steklov Inst. Math.* MR0871856

[3] Bártfai, P. (1966). Die Bestimmung der zu einem wiederkehrenden Prozess gehörenden Verteilungsfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation. *Studia Sci. Math. Hungar* **1** 161–168. MR0215377

[4] Berkes, I. and Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7** 29–54. MR0515811

[5] Bolthausen, E. (1982). Exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **10** 672–688. MR0659537

[6] Bretagnolle, J. and Massart, P. (1989). Hungarian constructions from the nonasymptotic view point. *Ann. Probab.* **17** 239–256. MR0972783

[7] Brown, L.D., Carter, A.V., Low, M.G. and Zhang, C.-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *Ann. Statist.* **32** 2074–2097. MR2102503

[8] Cai, T. T. and Zhou, H. H. (2009). Asymptotic Equivalence and Adaptive Estimation for Robust Nonparametric Regression. *Ann. Statist.* **37** 3204–3235. MR2549558

[9] Carter, A. V. and Pollard, D. (2004). Tusnády's inequality revisited. *Ann. Statist.,* **32** 2731–2741. MR2154001

[10] Chatterjee, S. (2012). A new approach to strong embeddings. *Probability Theory and Related Fields* **152** 231–264. MR2875758

[11] Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*, John Wiley & Sons, Chichester etc. MR1215046

[12] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*, Academic, New York. MR0666546

[13] Csörgő, M., Csörgő, S., Horváth, L. and Mason, D. M. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31–85. MR0815960

[14] de la Peña, V. H., Lai, T. L. and Shao, Q-M. (2009). *Self-normalized Processes. Limit Theory and Statistical Applications. Probability and its Applications* (New York). Springer-Verlag, Berlin. MR2488094

[15] Donoho, D. L. and Johnstone, I. M. (1995). Adapt to unknown smoothness via wavelet shrinkage. *J. Amer. Stat. Assoc.* **90** 1200–1224. MR1379464

[16] Donoho, D. L. and Yu , T. P.-Y. (2000). Nonlinear Pyramid Transforms Based on Median-Interpolation. *SIAM Journal of Math. Anal.* **31** 1030–1061. MR1759198

[17] Dudley, R. M. (2000). An exposition of Bretagnolle and Massart's proof of the KMT theorem for the uniform empirical process. In: Notes on empirical processes, lectures notes for a course given at Aarhus University, Denmark, August 2000.

[18] Einmahl, U. (1986). A refinement of the KMT-inequality for partial sum strong approximation. Technical Report Series of the Laboratory for Research in Statistics and Probability, Carleton University-University of Ottawa, No. 88.

[19] Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68. MR0996984

[20] Giné, E., Götze, F. and Mason, D.M. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* **25** 1514–1531. MR1457629

[21] Golubev, G. K., Nussbaum, M. and Zhou, H. H. (2010). Asymptotic equivalence of spectral density estimation and Gaussian white noise. *Ann. Statist.* **38** 181–214. MR2589320

[22] Grama, I. and Haeusler, E. (2000). Large deviations for martingales via Cramér's method. *Stochastic Process. Appl.* **85** 279–293. MR1731027

[23] Grama, I. and Nussbaum, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theory Relat. Fields* **111** 167–214 MR1633574

[24] Grama, I. and Nussbaum, M. (2002a). Asymptotic equivalence for nonparametric regression. *Mathematical Methods of Statistics*, **11** 11–36. MR1900972

[25] Grama, I. and Nussbaum, M. (2002b). A functional Hungarian construction for sums of independent random variables. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 923–957. MR1955345

[26] Hall, P. and Patil, P. (1996). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by wavelet methods, *J. Roy. Statist. Soc. Ser. B*, **58** 361–377. MR1377838

[27] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

[28] Hu, Z., Robinson, J. and Wang, Q. (2007). Cramér-type large deviations for samples from a finite population. *Ann. Statist.* **35** 673–696. MR2336863

[29] Jing, B.-Y, Shao, Q-M. and Wang, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31** 2167–2215. MR2016616

[30] Koltchinskii, V. (1994). Komlós-Major-Tusnády approximation for the general empirical process and Haar expansions of classes of functions. *J. Theoret. Probab.* **7** 73–118. MR1256392

[31] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample df. I Z. *Wahrsch. verw. Gebiete* **32** 111–131. MR0375412

[32] Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent rv's and the sample df. II Z. *Wahrsch. verw. Gebiete* **34** 33–58. MR0402883

[33] Lawler, G, F. and Trujillo Ferreras, J. A. (2007). Random walk loop soup. *Trans. Amer. Math. Soc.* **359** 767–787. MR2255196

[34] Major, P. (1976). The approximation of partial sums of independent RV's. Z. *Wahrsch. verw. Gebiete* **35** 213–220. MR0415743

[35] Major, P. (1978). On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.* **8** 487–517. MR0520959

[36] Major, P. (1999). The approximation of the normalized empirical ditribution function by a Brownian bridge. Technical report, Mathematical Institute of the Hungarian Academy of Sciences. Notes available from `http://www.renyi.hu/~major/probability/empir.html`.

[37] Mason, D. M. (2001). Notes on the KMT Brownian bridge approximation to the uniform empirical process. *In Asymptotic Methods in Probability and Statistics with Applications* (N. Balakrishnan, I. A. Ibragimov and V. B. Nevzorov, eds.) 351–369. Birkhäuser, Boston. MR1890338

[38] Mason, D. M. and van Zwet, W. R. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884 MR0893903

[39] Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069

[40] Massart, P. (2002). Tusnády's lemma, 24 years later. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 991–1007. MR1955348

[41] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430 MR1425959

[42] Petrov, V. V. (1975). *Sums of Independent Random Variables.* Springer-Verlag. (English translation from 1972 Russian edition). MR0388499

[43] Petrov, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Clarendon Press Oxford 1995 MR1353441

[44] Richter, H. *Über das dyadische Einbettungsschema von Komlós-Major-Tusnády*. Diplomarbeit, Ruhr-Universitaet Bochum, 1978; Referent: P. Gänssler, 1978

[45] Sakhanenko, A. I. (1982). On unimprovable estimates of the rate of convergence in invariance principle. *Colloquia Math. Soc. János Bolyai, 32 II. Nonparametric Statistical Inference, 779–783. Ed. by Gdnenko, B.V., Vincze, I. and Puri, M.L; North Holland, Amsterdam.* MR0669045

[46] Sakhanenko, A. I. (1985a). Convergence rate in the invariance principle for non-identically distributed variables with exponential moments. In:

Advances in Probability Theory: Limit Theorems for Sums of Random Variables (A. A. Borovkov, ed.) 2–73. Springer, New York.

[47] Sakhanenko, A. I. (1985b). Estimates in the invariance principles, In: *Trudy Inst. Mat. SO AN SSSR,* vol. 5, Nauka, Novosibirsk, pp 27-44. (Russian) MR0821751

[48] Sakhanenko, A. I. (1991). On the accuracy of normal approximation in the invariance principle *Siberian Adv. Math.* **1** 58–91. MR1138005

[49] Sakhanenko, A. I. (1996). Estimates for the Accuracy of Coupling in the Central Limit Theorem. *Siberian Mathematical Journal* **37** 811–823. MR1643327

[50] Saulis, L. I. (1969). Asymptotic expansions of probabilities of large deviations (in Russian). Litovskii Matemat. Sbomik (Lietuvos. Matematikos Rinkinys) **9** 605–625. MR0264742

[51] Saulis, L. I. and Statulevicius, V.A. (1991). *Limit theorems for large deviations.* Kluwer Academic Publishers. MR1171883

[52] Shao, Q-M. (1995). Strong approximation theorems for independent random variables and their applications. *J. Multivariate Anal.* **52** 107-130. MR1325373

[53] Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. MR0838963

[54] Statulevičius, V. and Jakimavičius, D. (1988). Estimates of semiinvariants and centered moments of stochastic processes with mixing. I. *Lithuanian Mathematical Journal* (Translated from Lietuvos Matematikos Rinkinys) **28** 112–129. MR0949647

[55] Tusnády, G. (1977). *A Study of Statistical Hypotheses*. Dissertation, The Hungarian Academy of Sciences, Budapest. (In Hungarian.)

[56] Wolf, W. (1977). Asymptotische Entwicklungen für Wahrscheinlichkeiten grosser Abweichungen. *Z. Wahrsch. verw. Geb.* **40** 239–256. MR0455089

[57] Zhou, H. H. (2004). *Minimax Estimation with Thresholding and Asymptotic Equivalence for Gaussian Variance Regression.* Ph.D. Dissertation. Cornell University, Ithaca, NY. MR2706282