A TALE OF TWO INEQUALITIES

David M. Mason

University of Delaware

ABSTRACT

In the first part of my talk I will show how an extension of a quantile inequality due to Komlós, Major and Tusnády (1975) yields a number of interesting couplings of a statistic and a standard normal random variable.

Associated with these couplings are certain generalized Bernstein-type inequalities. In order to apply these couplings it is often helpful to have a maximal Bernstein-type inequality.

This need led to a new and unexpected maximal Bernstein-type inequality, which will be described in the second part of my talk, along with applications.

This part of my talk is based upon joint work with Peter Kevei.

KMT (1975)

Komlós, Major and Tusnády (1975) proved the following powerful Gaussian coupling to partial sums of i.i.d. random variables.

Theorem [PS] *Let X be a random variable with mean 0 and variance $0 < \sigma^2 < \infty$. Also assume that its moment generating function $E\exp(hX)$ is finite for all h in a neighborhood of 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 $n \geq 1$,*

$$P\left\{\max_{1\leq k\leq n}\left|\sigma^{-1}\sum_{i=1}^{k}X_i - \sum_{i=1}^{k}Z_i\right| > D\log n + x\right\}$$

$$\leq C\exp\left(-\lambda x\right).$$

QUANTILE FUNCTION

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

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

denote the cumulative distribution function of Y_n.

Its *inverse distribution function* or *quantile function* is defined by

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

A DISTRIBUTIONAL IDENTITY

Let Z denote a standard normal random variable and Φ be its distribution 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

$$H_n(\Phi(Z)) = Y_n. \qquad \text{(HY)}$$

SPECIAL CASE

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

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

where $X_1, X_2, \ldots,$ are i.i.d. X satisfying the conditions of Theorem [PS].

QUANTILE INEQUALITY

Fundamental to the proof of Theorem [PS] is the following quantile inequality.

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* (SUM) *and* (HY), *and*

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

we have

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

COUPLING

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

$$P\left\{\sqrt{n}\,|Y_n - Z| > z\right\} \leq C \exp\left(-\lambda z\right),$$

BERKES-PHILIPP LEMMA

Another way of saying this is that via Lemma A1 of Berkes and Philipp (1979):

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 $z \geq 0$

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

BASIC QUANTILE INEQUALITY

The following quantile inequality is essentially due to KMT (1975) 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 by experts.

Let $\{F_n\}_{n \geq 1}$ be a sequence of cumulative distribution functions, not necessarily being that of a sequence of sums of i.i.d. random variables, and let Y_n be defined through the quantile function as in HY.

BASIC THEOREM *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* $0 < z \leq \varepsilon_n \sqrt{n}$

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

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

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

and

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

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

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

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

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

Specializing to $K_n = Ln^{1/2-1/p}$ we get:

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

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

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

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

and

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

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

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

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

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

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

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

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

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

Example 1 (Partial sums).

Let $X_1, X_2, ...,$ be i.i.d. X having mean zero and variance $0 < \sigma^2 < \infty$. Assume that for some $C > 0$, $D > 0$ and $0 < \beta \leq 1$, for all $x \geq 0$

$$P\left\{|X| > x\right\} \leq D \exp\left(-Cx^{\beta}\right).$$

By a Theorem in Saulis and Statulevičius (1991), the sequence of random variables

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

satisfies the conditions of the Corollaries 1 and 2, with

$$\beta = 4/(p+2),\ p \geq 2.$$

The case $p = 2$, i.e. $\beta = 1$ corresponds to the conditions of Theorem [PS].

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 (2003) shows that the assumptions of Corollary 1 hold with $p = 2$.

GGM Inequality

Now since (obviously)

$$Y_n = O_p(1),$$

we can apply Theorem 2.5 of Giné, Goetze and Mason (1997) to show that for suitable constants $b > 0$ and $c > 0$, for all $z \geq 0$ and $n \geq 1$

$$P\{|Y_n| \geq z\} \leq c \exp\left(-bz^2\right).$$

Thus we conclude by Corollary 2 that for positive constants C and λ, for all $z \geq 0$ and $n \geq 1$,

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

REMARK

Recall in Example 1 that for this last coupling inequality to hold for an un-self-normalized sum $S_n/(\sigma\sqrt{n})$ we required that X have a finite moment generating function in a neighborhood of zero.

Example 2 shows that self-normalizing dramatically reduces the assumptions needed for this coupling inequality to be valid.

Example 3 (dependent sums).

Let $X_1, X_2, \ldots,$ be a stationary sequence of random variables satisfying

$$EX_1 = 0 \text{ and } VarX_1 = 1.$$

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

$$Y_n =_d S_n/B_n.$$

The following examples are taken from Statulevičius and Jakimavičius (1988).

SOME SIGMA FIELDS

Let $\left\{\mathcal{F}, \mathcal{F}_s^t : 1 \leq s \leq t < \infty\right\}$ be a family of sigma fields 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) $\sigma\left\{X_u, 1 \leq s \leq u \leq t < \infty\right\} \subset \mathcal{F}_s^t$.

SOME MIXING CONDITIONS

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

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

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

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

In these last two expressions it is understood that $0/0 := 0$, whenever its occurs.

SOME MIXING RATES

For some $B > 0$ and $\mu > 0$

(M.1) $\alpha(s,t) \le Be^{-\mu(t-s)}$,

(M.2) $\varphi(s,t) \le Be^{-\mu(t-s)}$,

(M.3) $\psi(s,t) \le Be^{-\mu(t-s)}$.

Some Bounding Conditions:

(B.1) $|X_1| \le C$ for some $0 < C < \infty$,

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

Applying results in Statulevičius and Jakimavičius (1988) we get that Corollaries 1 and 2 hold:

Under conditions (M.1) and (B.1), with $p = 6$;

under conditions (M.1) and (B.2), with $p = 10$;

under $m-$dependence and (B.1), with $p = 2$;

under $m-$dependence and (B.2), with $p = 6$.

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

Under conditions (M.2) and (B.1), with $p = 2$;

under conditions (M.2) and (B.2), with $p = 6$;

under conditions (M.3) and (B.2), with $p = 2$.

Specializing to m-dependent sums

Form the partial sum process:

$$Y_n\left(\frac{i}{n}\right) = \frac{S_i}{B_n}, \text{ for } i = 0, 1, \ldots, n,$$

where $S_0 = 0$, and let for $i = 1, \ldots, n$, and for $(i-1)/n < t < i/n$

$$Y_n(t) = \left(S_{i-1} + n\left(t - \frac{i-1}{n}\right) X_i\right) / B_n, .$$

Further let $W(t)$, $0 \le t \le 1$, denote a standard Wiener process. Let P_{Y_n} and P_W denote the distributions on $\mathcal{C}[0,1]$. One can use the above results to show that for some constant $c > 0$

$$L\left(P_{Y_n}, P_W\right) \le \frac{c\sqrt{\log n}}{n^{1/4}},$$

where L denotes the Prokhorov distance.

APPROXIMATION INEQUALITY

There exist positive constants c_i, $i = 1, \ldots, 6$, such that for each $n \geq m$ a version of $Y_n(\cdot)$ and W can be defined on the same probability space such that for all $z \geq 0$

$$P\left\{\sup_{0\leq t\leq 1} |Y_n(t) - W(t)| > \frac{c_1 z + c_2\sqrt{\log n}}{n^{1/4}}\right\}$$

$$\leq c_3 \exp\left(\frac{-c_4 z^2}{c_5 + c_6\left(z/n^{1/4}\right)}\right).$$

STRONG APPROXIMATION

The above is not a strong approximation. Applying a special case of a result of Shao and Lu (1987) shows that in this setup there exist i.i.d. standard normal random variables $Z_1, Z_2, \ldots,$ on the same probability space as $X_1, X_2, \ldots,$ such that with $\mathbb{Z}_m = Z_1 + \cdots + Z_m, m \geq 1,$

$$\max_{1 \leq m \leq n} \left| \sigma^{-1} S_m - \mathbb{Z}_m \right| = O\left((\log n)^{9/4+\varepsilon} \, n^{1/4} \right), \text{ a.s.}$$

BERNSTEIN INEQUALITY

Let $X_1, X_2, \ldots,$ be a sequence of independent random variables such that for all $i \geq 1$, $EX_i = 0$ and for some $\kappa > 0$ and $v > 0$ for integers $m \geq 2$, $E|X_i|^m \leq vm!\kappa^{m-2}/2$. The classic Bernstein inequality (cf. p. 855 of Shorack and Wellner (1986) says that in this situation for all $n \geq 1$ and $t \geq 0$

$$\mathbf{P}\left\{\left|\sum_{i=1}^{n} X_i\right| > t\right\} \leq 2\exp\left\{-\frac{t^2}{2vn + 2\kappa t}\right\}.$$

Moreover, (cf. Theorémè B.2 in Rio (200) its maximal form also holds, i.e. we have

$$\mathbf{P}\left\{\max_{1\leq j\leq n}\left|\sum_{i=1}^{j} X_i\right| > t\right\} \leq 2\exp\left\{-\frac{t^2}{2vn + 2\kappa t}\right\}.$$

GENERAL BERNSTEIN INEQUALITY

It turns out that, under a variety of assumptions, a sequence of not necessarily independent random variables $X_1, X_2, \ldots$, will satisfy a generalized Bernstein-type inequality of the following form: for suitable constants $A > 0$, $a > 0$, $b \geq 0$ and $0 < \gamma < 2$ for all $i \geq 0$, $n \geq 1$ and $t \geq 0$,

$$\mathbf{P}\{|S(i+1, i+n)| > t\} \leq A \exp\left\{-\frac{at^2}{n + bt^\gamma}\right\}, \tag{GB}$$

where for any choice of $1 \leq i \leq j < \infty$ we denote the partial sum $S(i,j) = \sum_{k=i}^{j} X_k$. Here are some examples.

Bernstein Example 1

Let $X_1, X_2, \ldots,$ be a stationary sequence satisfying

$$EX_1 = 0 \text{ and } VarX_1 = 1.$$

For each integer $n \geq 1$ 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 $n \geq 1$.

Statulevičius and Jakimavičius (1988) prove that the partial sums satisfy GB with constants depending on the particular mixing and bounding condition that the sequence may fulfill.

BENTKUS AND RUDZKIS

Their Bernstein-type inequalities are derived via the following result of Bentkus and Rudzkis (1980) relating cumulant behavior to tail behavior:

For an arbitrary random variable ξ with expectation 0, whenever there exist $\gamma \geq 0$, $H > 0$ and $\Delta > 0$ such that its cumulants $\Gamma_k(\xi)$ satisfy $|\Gamma_k(\xi)| \leq (k!/2)^{1+\gamma} H/\Delta^{k-2}$ for $k = 2, 3, \ldots$, then for all $x \geq 0$

$$\mathbf{P}\{\pm\xi > x\}$$

$$\leq \exp\left\{-\frac{x^2}{2\left(H + \left(x/\Delta^{1/(1+2\gamma)}\right)^{(1+2\gamma)/(1+\gamma)}\right)}\right\}.$$

Bernstein Example 2

Doukhan and Neumann (2007) have shown using the result in Bentkus and Rudzkis (1980) cited in the previous example that if a sequence of mean zero random variables $X_1, X_2, \ldots,$ satisfies a general covariance condition then the partial sums satisfy GB.

Refer to their Theorem 1 and Remark 2, and also see Kallabis and Neumann (2006).

Bernstein Example 3

Assume that $X_1, X_2, \ldots,$ is a strong mixing sequence with mixing coefficients $\alpha(n)$, $n \geq 1$, satisfying for some $c > 0$, $\alpha(n) \leq \exp(-2cn)$. Also assume that $EX_i = 0$ for some $M > 0$ $|X_i| \leq M$, for all $i \geq 1$. Theorem 2 of Merlevéde, Peligrad and Rio (2009) implies that for some constant $C > 0$ for all $t \geq 0$ and $n \geq 1$,

$$\mathbf{P}\{|S_n| \geq t\} \leq \exp\left(-\frac{Ct^2}{nv^2 + M^2 + tM(\log n)^2}\right),$$

where $S_n = \sum_{i=1}^n X_i$ and

$$v^2 = \sup_{i>0}\left(Var(X_i) + 2\sum_{j>i}|cov(X_i, X_j)|\right).$$

EXPLANATION

To see how this last example satisfies GB, notice that for any $0 < \eta < 1$ there exists a $D_1 > 0$ such that for all $t \geq 0$ and $n \geq 1$,

$$nv^2 + M^2 + tM(\log n)^2 \leq n\left(v^2 + M^2\right) + D_2 t^{1+\eta}.$$

Thus GB holds with $\gamma = 1 + \eta$ for suitable $A > 0$, $a > 0$ and $b \geq 0$.

GENERAL MAXIMAL BERNSTEIN INEQUALITY

For any choice of $1 \le i \le j < \infty$ define

$$M(i,j) = \max\{|S(i,i)|, \ldots, |S(i,j)|\}.$$

Somewhat unexpectedly, if a sequence of random variables $X_1, X_2, \ldots,$ satisfies a Bernstein-type inequality of the form GB, then without any additional assumptions a modified version of it also holds for $M(m+1, m+n)$.

GMB Inequality *Assume that for constants $A > 0$, $a > 0$, $b \ge 0$ and $\gamma \in (0,2)$, inequality* GB *holds for all $i \ge 0, n \ge 1$ and $t \ge 0$. Then there exist constants $c > 0$ and $C > 0$ depending only on A, a, b and γ such that for all $m \ge 0$, $n \ge 1$ and $t \ge 0$,*

$$\mathbf{P}\{M(m+1, m+n) > t\} \le C \exp\left\{-\frac{ct^2}{n + bt^\gamma}\right\}.$$

Moreover, if $0 < \gamma \le 1$ we can take $c = a$ and if $1 < \gamma < 2$, $c < a$ can be chosen arbitrarily close to a.

MOTIVATION

The GMB inequality was motivated by Theorem 2.2 of Móricz, Serfling and Stout (1982), who showed that whenever for a suitable positive function $g(i,j)$ of $(i,j) \in \{1,2,\dots\} \times \{1,2,\dots\}$, positive function $\phi(t)$ defined on $(0,\infty)$ and constant $K > 0$, for all $1 \le i \le j < \infty$ and $t > 0$,

$$\mathbf{P}\{|S(i,j)| > t\} \le K \exp\{-\phi(t)/g(i,j)\},$$

then there exist constants $c > 0$ and $C > 0$ such that for all $m \ge 0$, $n \ge 1$ and $t > 0$,

$$\mathbf{P}\{M(m+1, m+n) > t\} \le C \exp\{-c\phi(t)/g(1,n)\}.$$

This inequality is clearly not applicable to obtain a maximal form of the generalized Bernstein inequality.

APPLICATIONS OF GMB INEQUALITY

An obvious application of the GMB inequality is the following bounded law of the iterated logarithm.

Bounded LIL *Under the assumptions of the previous theorem, with probability* 1,

$$\limsup_{n\to\infty} \frac{|S(1,n)|}{\sqrt{n \log\log n}} \leq \frac{1}{\sqrt{a}}.$$

OBSERVATION

In general one cannot replace "$\leq$ " by "$=$" our bounded LIL. To see this, let Y, $Z_1, Z_2, \dots$ be a sequence of independent random variables such that Y takes on the value 0 or 1 with probability 1/2 and $Z_1, Z_2, \dots$ are independent standard normals. Now define $X_i = YZ_i$, $i = 1, 2, \dots$ It is easily checked that assumptions of the GMB inequality are satisfied with $A = 2$, $a = 1/2$, $b = 0$ and $\gamma = 1$.

When $Y = 1$ the usual law of the iterated logarithm gives with probability 1,

$$\limsup_{n\to\infty} |S(1,n)|/\sqrt{n \log\log n} = \sqrt{2} = 1/\sqrt{a}$$

whereas, when $Y = 0$ the above limsup is 0. This agrees with the bounded LIL, which says that with probability 1 the limsup is $\leq \sqrt{2}$.

However, we see that with probability 1/2 it equals $\sqrt{2}$ and with probability 1/2 it equals 0.

The GMB inequality is also a useful tool in establishing approximation inequalities for stationary partial sum processes under the mixing and bounding conditions that were defined above of the sort that was described here for the stationary m-dependent partial sum process.