Berry-Esseen bounds for general nonlinear statistics, with applications to Pearson’s and non-central Student’s and Hotelling’s

Iosif Pinelis\textsuperscript{1,2} and Raymond Molzon\textsuperscript{1}

\textsuperscript{1}Department of Mathematical Sciences  
Michigan Technological University  
\textsuperscript{2}Supported in part by NSF grant DMS-0805946

September 18, 2011
Main results

Supporting and/or related results
Initial motivation: Pitman’s ARE between Pearson’s, Kendall’s, and Spearman’s correlation coeffs.

- Needed: closeness to normality uniformly near $H_0$.
- Kendall’s and Spearman’s coeffs. are $U$-statistics, with known BE bounds.
- For Pearson’s, a BE bound is not found in literature. Hardly surprising:
  - an optimal BE bound for Student’s $t$: obtained only in ’96, by Bentkus and Götze;
  - A necessary and sufficient condition, in the i.i.d. case, for $t$ to be asymptotically normal: obtained only in ’97, by Giné, Götze and Mason.
Pearson’s $R$: general idea

Let $(Y, Z), (Y_1, Z_1), \ldots, (Y_n, Z_n)$ be i.i.d. random points in $\mathbb{R}^2$; w.l.o.g. $E Y = E Z = 0$ and $E Y^2 = E Z^2 = 1$.

Pearson’s $R := \frac{\overline{YZ} - \overline{Y} \overline{Z}}{\sqrt{\overline{Y^2} - \overline{Y}^2} \sqrt{\overline{Z^2} - \overline{Z}^2}} = f(\overline{V})$,

where $\overline{V} := \frac{1}{n} \sum_1^n V_i$ and the $V_i$’s are iid copies of $V := (Y, Z, Y^2 - 1, Z^2 - 1, YZ - \rho)$,

with $\rho := E YZ = \text{Corr}(Y, Z)$, so that $E V = 0$.

So, $f(\overline{V}) \approx f(0) + L(\overline{V})$, where $L := f'(0)$ is a linear functional.

From here, using exp. ineqs. for sums in $B$-spaces by Pinelis–Sakhanenko ’85: BE bounds $O(n^{-1/2} \ln^{3/2} n)$ if $\|V\|_3 < \infty$, and $O(n^{1-p/2})$ if $\|V\|_p < \infty$ for some $p \in (2, 3)$. 
Given:

- an abstract nonlinear statistic $T$;
- an abstract linear statistic $W$.

For $\Delta := T - W$, start with

$$-\mathbb{P}(z - |\Delta| < W \leq z) \leq \mathbb{P}(T \leq z) - \mathbb{P}(W \leq z) \leq \mathbb{P}(z < W \leq z + |\Delta|).$$

A number of applications were given by Chen and Shao. We modify their method, apply it to $f(\overline{V}) \approx f(0) + L(\overline{V})$, and use other tools to get:

BE-type uniform and nonuniform bounds for statistics of the form $f(\overline{V})$;

in fact, for non-identically distributed random vectors as well.
A resulting corollary for Pearson’s $R$:

If e.g. $\rho = EYZ = 0$ and $\tilde{\sigma} := \sqrt{E X^2 Y^2} \neq 0$ then

$$\left| \mathbb{P}\left( \frac{R}{\tilde{\sigma}/\sqrt{n}} \leq z \right) - \Phi(z) \right| \leq \frac{6.75}{\sqrt{n}} \left( \|Y\|_6^6 + \|Z\|_6^6 \right) \left(1 + \tilde{\sigma}^{-3}\right).$$

Recall:

here $V = (Y, Z, Y^2 - 1, Z^2 - 1, YZ)$;

so, $\|V\|_3^3 \simeq \|Y\|_6^6 + \|Z\|_6^6$.

Cf. Chibisov '80: Asymptotic expansion for the distr. of statistics admitting a stochastic expansion.
Let $Y, Y_1, \ldots, Y_n$ be iid with $E Y = 0$ and $E Y^2 = 1$. Let

$$T := \frac{Y_1 + \cdots + Y_n}{\sqrt{Y_1^2 + \cdots + Y_n^2}} = \frac{\sqrt{n} \bar{Y}}{\sqrt{\sum Y_i^2}}.$$ 

Shao '05:

$$|\mathbb{P}(T \leq z) - \Phi(z)| \leq 25 \frac{E |Y|^3}{\sqrt{n}} I\{|Y| \leq \frac{\sqrt{n}}{2}\} + 10.2 E Y^2 I\{|Y| > \frac{\sqrt{n}}{2}\}$$

$$\leq 25 \frac{\|Y\|^3}{\sqrt{n}},$$

earlier, Bentkus and Götze '96: same without explicit constants.
Central $t$ and self-normalized sums: Chibisov '79–80; Slavova '85; Novak '00 and '05; Nagaev '02; and Pinelis '11 – an ad hoc method

Pinelis '11:

$$\left| \mathbb{P}(T \leq z) - \Phi(z) \right| \leq \frac{1}{\sqrt{n}} \left( A_3 \|Y\|^3_3 + A_4 \|Y^2 - 1\|_2 + A_6 \frac{\|Y^2 - 1\|^3_3}{\|Y\|^3_3} \right)$$

for $(A_3, A_4, A_6) \in \{(1.53, 1.52, 1.34), (10.94, 9.40, 11.06 \times 10^{-6})\}$; the constants are slightly worse without the iid assumption.

Especially after truncation, this compares favorably with Shao’s result.
Pinelis and Molzon ’11, based on a modification of the Chen–Shao result for abstract nonlinear statistics:

\[ |\mathbb{P}(T \leq z) - \Phi(z)| \leq \frac{3.68\|Y\|_3^3 + 2.60\|Y\|_4^4 - 0.98}{\sqrt{n}}. \]

Pinelis and Molzon ’11, based on a general result for statistics of form \( f(\overline{V}) \):

\[ |\mathbb{P}(T \leq z) - \Phi(z)| \leq \frac{5.08\|Y\|_3^3 + 5.08\|Y\|_4^6 + 0.26\|Y^2 - 1\|_3^3}{\sqrt{n}}. \]
Self-normalized sums, nonuniform BE bounds

\[ |P(T \leq z) - \Phi(z)| \leq \frac{g(z)}{\sqrt{n}} \left( A_3 \| Y \|_3^3 + A_4 \| Y \|_4^8 + A_6 \| Y^2 - 1 \|_3^3 \right), \]

where \( g(z) := \frac{1}{z^3} + \frac{w_g}{e^{z/2}}, \quad z \in (0, \omega \sqrt{n}] \),

\( w_g \in \{0, 1\}, \quad \omega \in \{0.1, 0.5\} \), and \((A_3, A_4, A_6)\) as follows:

<table>
<thead>
<tr>
<th>( w_g = 1 )</th>
<th>( \omega = 0.1 )</th>
<th>( \omega = 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w_g = 0 )</td>
<td>( \omega = 0.1 )</td>
<td>( \omega = 0.5 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( A_3 )</th>
<th>( A_4 )</th>
<th>( A_6 )</th>
<th>( A_3 )</th>
<th>( A_4 )</th>
<th>( A_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>42</td>
<td>41</td>
<td>59</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>54</td>
<td>27</td>
<td>11</td>
<td>122</td>
<td>61</td>
<td>25</td>
</tr>
<tr>
<td>190</td>
<td>189</td>
<td>159</td>
<td>260</td>
<td>259</td>
<td>251</td>
</tr>
<tr>
<td>258</td>
<td>127</td>
<td>52</td>
<td>399</td>
<td>197</td>
<td>80</td>
</tr>
</tbody>
</table>

Comment: \( \max_{z > 0} \frac{e^{-z/2}}{1/z^3} \approx 10.75 \), attained at \( z = 6 \).

Cf.: best known nonuniform BE bound for sums of i.i.d. r.v.'s (Michel '81 cum Tyurin '09): \( 30.225 \frac{\| Y \|_3^3}{(|z|^3 + 1)^{1/2} \sqrt{n}} \).
Self-normalized sums vs. Student’s statistic

For all $n > 1$

$$\left| \sup_{z \in \mathbb{R}} \mathbb{P}(t \leq z) - \Phi(z) \right| - \sup_{z \in \mathbb{R}} \left| \mathbb{P}(T \leq z) - \Phi(z) \right| < \frac{C}{n-1},$$

where $t$ is the Student statistic, $T$ is again the self-normalized sum,

$$C := \left( k - \frac{1}{2} \right) e^{-k} \sqrt{\frac{k}{\pi}} = 0.162 \ldots, \quad \text{and} \quad k := 1 + \frac{\sqrt{3}}{2}.$$
Exact bounds on the closeness between $t_p$ and $t_\infty = N(0, 1)$

Let $d_{TV}(p)$ and $d_{Ko}(p)$ denote, resp., the total-variation and Kolmogorov distances between $t_p$ and $t_\infty = N(0, 1)$. Then

$$\frac{1}{2} d_{TV}(p) = d_{Ko}(p) < \frac{C}{p} \quad \forall p \in [4, \infty),$$

where

$$C := \frac{1}{4} \sqrt{\frac{7 + 5\sqrt{2}}{\pi e^{1+\sqrt{2}}}} = \lim_{p \to \infty} p d_{Ko}(p) = 0.158 \ldots,$$

so that $C$ is the best possible factor.
Let $G_p$ stand for the tail function of $t_p$. Then

$$0 < p < q \leq \infty \implies \frac{G_q(x)}{G_p(x)} \text{ is (strictly) decr. in } x \geq 0,$$

whence the stochastic majorization:

$$G_q(x) < G_p(x) \quad \forall x > 0.$$
For each \( h > 0 \) and each \( w \in \mathbb{R} \), the maximum of the tilted mean
\[
\frac{E X e^{h(X \wedge w)}}{E e^{h(X \wedge w)}}
\]
given \( E X \) and \( E X^2 \) is attained when \( X \) has a two-point distribution.

For \( E X = 0 \) and \( w > 0 \), this maximum is
\[
< \frac{e^{hw} - 1}{w} E X^2,
\]
and the factor \( \frac{e^{hw} - 1}{w} \) is the best possible.
For each $h > 0$ and each $w > 0$, the minimum of

$$E e^{h(X \wedge w)}$$

given $E X \geq 0$ and $E X^2$ is attained when $X$ has a two-point distribution.

For each $w > 0$, the minimum of these minima over all $h > 0$ is strictly positive (not so if $X \wedge w$ is replaced by $X I\{X \leq w\}$).
An asymptotically Gaussian bound on the Rademacher tails

\[ \mathbb{P}(a_1 \varepsilon_1 + \cdots + a_n \varepsilon_n \geq x) \leq \mathbb{P}(Z > x) + \frac{C \varphi(x)}{9 + x^2} \]

\[ < \mathbb{P}(Z > x) \left(1 + \frac{C}{x}\right) \quad \forall x > 0, \]

where \( \varepsilon_1, \ldots, \varepsilon_n \) are independent Rademacher r.v.’s,

\[ a_1^2 + \cdots + a_n^2 = 1, \]

\[ Z \sim N(0, 1) \text{ with density } \varphi, \text{ and } \]

\[ C := 5\sqrt{2\pi}e \mathbb{P}(|Z| < 1) = 14.10 \ldots \text{ is a best possible constant factor: the 1st inequality above turns into the equality when } x = n = 1. \]
Refined and generalized Bennett-Hoeffding bound

Let $X_1, \ldots, X_n$ be independent r.v.'s, with $S := X_1 + \cdots + X_n$. Take any $\sigma, y, \beta$ in $(0, \infty)$ s.t. $\varepsilon := \frac{\beta}{\sigma^2 y} \in (0, 1)$. Suppose that

$$
\sum_i E X_i^2 \leq \sigma^2, \quad \sum_i E(X_i)^3 \leq \beta, \quad E X_i \leq 0, \quad \text{and } X_i \leq y,
$$

for all $i$. Then

$$
E f(S) \leq E f(\eta_{\varepsilon, \sigma, y})
$$

for all $f \in C^2$ s.t. $f$ and $f''$ are nondecreasing and convex, where

$$
\eta_{\varepsilon, \sigma, y} := \Gamma (1-\varepsilon) \sigma^2 + y \tilde{\Pi}_{\varepsilon \sigma^2 / y^2},
$$

$\Gamma_{a^2} \sim N(0, a^2)$, $\tilde{\Pi}_\theta := \Pi_\theta - \theta$, $\Pi_\theta \sim \text{Poisson}(\theta)$, and $\Gamma_{a^2}$ and $\Pi_\theta$ are independent.

Corollary: $\mathbb{P}(S \geq x) \leq \frac{2e^3}{9} \mathbb{P}^{LC}(\eta_{\varepsilon, \sigma, y} \geq x) \quad \forall x \in \mathbb{R},$

where $\mathbb{R} \ni x \mapsto \mathbb{P}^{LC}(\eta \geq x)$ is the least log-concave majorant of $\mathbb{R} \ni x \mapsto \mathbb{P}(\eta \geq x)$. 
Positive-part moments via the Fourier–Laplace transform

For $p$ and $s$ in $(0, \infty)$, $j \in \{-1, \lceil p - 1 \rceil\}$, and any r.v. $X$ with $E e^{sX} < \infty$ one has

$$E X^p_+ = \frac{\Gamma(p + 1)}{\pi} \int_0^\infty \Re \frac{E e^j((s + it)X)}{(s + it)^{p+1}} \, dt,$$

where $e_j(z) := e^z - \sum_{r=0}^j \frac{z^r}{r!}$, with $e_{-1}(z) \equiv e^z$.

Moreover, if $E |X|^p < \infty$ then

$$E X^p_+ = \frac{E X^p}{2} \mathbf{1}\{p \in \mathbb{N}\} + \frac{\Gamma(p + 1)}{\pi} \int_0^\infty \Re \frac{E e_{\ell}(itX)}{(it)^{p+1}} \, dt.$$
\[ \mathcal{F}^{1,2}_+ := \{ f \in C^1(\mathbb{R}) : f(0) = 0, \; f \text{ is even}, \]
\[ f' \text{ is nondecreasing and concave on } [0, \infty) \}; \]
e.g., \( | \cdot |^p \in \mathcal{F}^{1,2}_+ \) \( \forall p \in (1, 2] \). Then
\[ E f(S_n) \leq E f(X_1) + C_f \sum_{j=2}^n E f(X_j) \]
where \( f \in \mathcal{F}^{1,2}_+ \setminus \{0\} \), \((S_j)_{j=1}^n\) is a martingale, and the constant \( C_f := \sup_{0 < x < s < \infty} \frac{1}{f(s)} (f(x - s) - f(x) + sf'(x)) \) is the best possible, for each \( f \in \mathcal{F}^{1,2}_+ \setminus \{0\} \); also, \( \{ C_f : f \in \mathcal{F}^{1,2}_+ \setminus \{0\} \} = [1, 2] \), and \( C_f = 1 \) if \( f(x) \equiv x^2 \).
Corollary: concentration of measure for separately Lipshitz (sep-Lip) functions

Let $X_1, \ldots, X_n$ be independent r.v.'s with values in $\mathcal{X}_1, \ldots, \mathcal{X}_n$, resp.; here, all spaces and functions are measurable. Suppose

$$Y := g(X_1, \ldots, X_n)$$

and the sep-Lip condition

$$|E g(x_1, \ldots, x_{i-1}, \tilde{x}_i, X_{i+1}, \ldots, X_n) - E g(x_1, \ldots, x_{i-1}, x_i, X_{i+1}, \ldots, X_n)| \leq \rho_i(\tilde{x}_i, x_i)$$

holds for some functions $\rho_i : \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}$ and all $i$, $x_j \in \mathcal{X}_j$, and $\tilde{x}_i \in \mathcal{X}_i$.

(The sep-Lip condition is easier to check and more generally applicable than joint-Lip. On the other hand, when joint-Lip holds, generally better bounds can be obtained.)
Then $\forall f \in \mathcal{F}_{+}^{1,2} \setminus \{0\}$ $\forall x_i \in \mathcal{X}_i$

$$E f(Y) \leq f(E Y) + \kappa_f C_f \sum_{i=1}^{n} E f(\rho_i(X_i, x_i)),$$ (*)

where

$$\kappa_f := \sup_{0 < a < c < s/2} \frac{c f(s - c) + (s - c)f(c)}{c f(s - c + a) + (s - c)f(a - c)}.$$ 

Moreover, $\{\kappa_f : f \in \mathcal{F}_{+}^{1,2} \setminus \{0\}\} = [1, 2]$.

Also, $\kappa_f = C_f = 1$ if $f(x) \equiv x^2$; for this $f$, inequality (*) in the case when $\mathcal{X}_1 = \cdots = \mathcal{X}_n$ is a B-space, $g(x_1, \ldots, x_n) = \|x_1 + \cdots + x_n\|$, and $\rho_i(\tilde{x}_i, x_i) = \|\tilde{x}_i - x_i\|$ was obtained by Pinelis '85.
Let $\mathcal{F}^{2,3}$ be the class (introduced by Figiel, Hitczenko, Johnson, Schechtman and Zinn ’97) of all $f : \mathbb{R} \to \mathbb{R}$ such that $f(0) = 0$, $f$ is even and convex, $[0, \infty) \ni t \mapsto f(\sqrt{t})$ is convex, and $f''$ is concave on $[0, \infty)$.

Then for any $f \in \mathcal{F}^{2,3}$, any zero-mean r.v. $X$, and any $t \in \mathbb{R}$

$$\mathbb{E} f(X) \leq c \mathbb{E} f(X + t),$$

where $c := \frac{17+7\sqrt{7}}{27} = 1.315\ldots$ is the best possible factor.
Thank you!