On the Bennett-Hoeffding inequality

Iosif Pinelis [1,2,3]

[1]Department of Mathematical Sciences
Michigan Technological University
[2]Supported by NSF grant DMS-0805946
[3]Paper available at `http://arxiv.org/abs/0902.4058`

June 1, 2009

Intro: Bennett-Hoeffding (BH) and other ineqs.:

$X_1, \dots, X_n$: indep. 0-mean real-valued r.v.'s s.t.
$X_i \leqslant y$ a.s. for some $y > 0$ and all i.
$S := X_1 + \dots + X_n$
$\sigma := \sqrt{\sum_i \mathsf{E}\, X_i^2} \in (0, \infty)$.
BH ineq.: $\forall x \geqslant 0$

$$\mathsf{P}(S \geqslant x) \leqslant \mathrm{BH}(x) := \mathrm{BH}_{\sigma^2, y}(x) := \exp\Big\{ -\frac{\sigma^2}{y^2}\, \psi\Big(\frac{xy}{\sigma^2}\Big)\Big\},$$

$$\psi(u) := (1+u)\ln(1+u) - u.$$

The BH ineq. has been generalized for X_i's not indep. and/or are not real-valued.

BH, continued

The BH ineq. is based on the optimal bound on the exp. moments:

$$\mathsf{E}\, e^{\lambda S} \leqslant \mathsf{BH}_{\exp}(\lambda) := \exp\left\{\frac{e^{\lambda y} - 1 - \lambda y}{y^2}\, \sigma^2\right\} \quad (\lambda > 0).$$

That is,

$$\mathsf{BH}(x) = \inf_{\lambda > 0} e^{-\lambda x}\, \mathsf{BH}_{\exp}(\lambda).$$

Several authors: attempts at refining BH by accounting for truncated pth moments ($p > 2$). However, in contrast with BH, their bounds were not the best possible in their own terms.

Pinelis-Utev (PU) bound

Best possible exp. bounds refining BH: by Pinelis and Utev '89:

$$\mathsf{E}\, e^{\lambda S} \leqslant \mathrm{PU}_{\exp}(\lambda) := \exp\left\{\frac{\lambda^2}{2}\,(1-\varepsilon)\sigma^2 + \frac{e^{\lambda y}-1-\lambda y}{y^2}\,\varepsilon\sigma^2\right\},$$

$$\varepsilon := \frac{\beta_3^+}{\sigma^2 y}, \quad \beta_3^+ := \sum_i \mathsf{E}(X_i)_+^3, \quad \lambda > 0;$$

$$\text{so}, \mathsf{P}(S \geqslant x) \leqslant \mathrm{PU}(x) := \inf_{\lambda>0} e^{-\lambda x}\,\mathrm{PU}_{\exp}(\lambda), \quad x \geqslant 0.$$

Note: $\varepsilon \in (0,1)$. Also, $\frac{\lambda^2}{2} < \frac{e^{\lambda y}-1-\lambda y}{y^2}$ for $\lambda > 0$ and $y > 0$; so, $\mathrm{PU}_{\exp}(\lambda) \leqslant \mathrm{BH}_{\exp}(\lambda)$ and $\mathrm{PU}(x) \leqslant \mathrm{BH}(x)$. Moreover, PU $<<$ BH if $\varepsilon << 1$, which is the case for i.i.d. X_i's with finite $\mathsf{E}\,X_i^2$ and $\mathsf{E}(X_i)_+^3$, when n is large and $y \asymp \sqrt{n}$ (as e.g. in proofs of non-uniform Berry-Esseen type bounds).

PU, continued

The exp. bounds $\mathrm{BH}_{\mathrm{exp}}(\lambda)$ and $\mathrm{PU}_{\mathrm{exp}}(\lambda)$ are exact:
$\mathrm{BH}_{\mathrm{exp}}(\lambda) = \sup \mathsf{E}\, e^{\lambda S}$ with λ, y, and σ fixed;
$\mathrm{PU}_{\mathrm{exp}}(\lambda) = \sup \mathsf{E}\, e^{\lambda S}$ with λ, y, σ, and ε fixed.

If $\varepsilon << 1$, then $\mathrm{PU}_{\mathrm{exp}}(\lambda) \approx e^{\lambda^2\sigma^2/2}$ and $\mathrm{PU}(x) \approx e^{-x^2/(2\sigma^2)}$.

However, even for $Z \sim \mathrm{N}(0,1)$, the best exp. bound $e^{-x^2/2}$ on $\mathsf{P}(Z \geqslant x)$ is “missing” a factor $\asymp \frac{1}{x}$ for large $x > 0$, since $\mathsf{P}(Z \geqslant x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$ as $x \to \infty$.
Cause of this deficiency: the class of all incr. exp. moment functs. is too small.

Towards eliminating the deficiency: richer classes of moment functs.

Definition

$\mathcal{H}_+^\alpha$:=the class of all functs. f s.t. $f(u)=\int_{-\infty}^{\infty}(u-t)_+^\alpha\,\mu(dt)$ for some Borel measure $\mu\geqslant 0$ and $\forall u\in\mathbb{R}$, where $\alpha>0$.

Note:

$$0\leqslant\beta<\alpha\quad\text{implies}\quad\mathcal{H}_+^\alpha\subseteq\mathcal{H}_+^\beta.$$

Proposition

For α natural: $f\in\mathcal{H}_+^\alpha$ iff $f^{(\alpha-1)}$ is convex and $f^{(j)}(-\infty+)=0$ for $j=0,1,\dots,\alpha-1$.

Eliminating the deficiency: optimal tail comparison

Theorem (special case of Th 3.11 in Pinelis '98)

Let $\alpha > 0$, ξ *and* η *be any r.v.'s s.t. the tail* $\mathsf{P}(\eta \geqslant u)$ *is log-concave in* $u \in \mathbb{R}$. *Then*

$$\mathsf{E}\, f(\xi) \leqslant \mathsf{E}\, f(\eta) \quad \textit{for all } f \in \mathcal{H}_+^\alpha$$

implies

$$\begin{aligned} \mathsf{P}(\xi \geqslant x) \leqslant P_\alpha(\eta; x) &:= \inf_{t \in (-\infty, x)} \frac{\mathsf{E}(\eta - t)_+^\alpha}{(x - t)^\alpha} \\ &\leqslant c_{\alpha,0}\, \mathsf{P}(\eta \geqslant x) \quad \forall x \in \mathbb{R}, \end{aligned}$$

where $c_{\alpha,0} := \Gamma(\alpha + 1)(e/\alpha)^\alpha$, *the best possible const. factor.*

A similar result for $\alpha = 1$: in Shorack and Wellner '86.

The least log-concave (LC) majorant of the tail

Definition

Let $\mathbb{R} \ni x \mapsto \mathsf{P}^{\mathsf{LC}}(\eta \geqslant x)$ be the least LC majorant of the tail funct. $\mathbb{R} \ni x \mapsto \mathsf{P}(\eta \geqslant x)$.

Remark

$\mathsf{P}^{\mathsf{LC}}(a + b\eta \geqslant x) = \mathsf{P}^{\mathsf{LC}}(\eta \geqslant \frac{x-a}{b})$ *for all* $x \in \mathbb{R}$, $a \in \mathbb{R}$, $b > 0$.

Remark

W/out the log-concavity of $\mathsf{P}(\eta \geqslant u)$ *in the comparison Theorem, one still has*

$$\mathsf{P}(\xi \geqslant x) \leqslant c_{\alpha,0}\, \mathsf{P}^{\mathsf{LC}}(\eta \geqslant x) \quad \forall x \in \mathbb{R}.$$

Back to the BH and PU exp. bounds: probabilistic interpretation

$$\mathrm{BH}_{\mathrm{exp}}(\lambda) = \mathsf{E}\exp\left\{\lambda y \tilde{\Pi}_{\sigma^2/y^2}\right\} \quad \text{and}$$
$$\mathrm{PU}_{\mathrm{exp}}(\lambda) = \mathsf{E}\exp\left\{\lambda\big(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}\big)\right\}$$

for all $\forall\lambda > 0$, where

$$\tilde{\Pi}_\theta := \Pi_\theta - \mathsf{E}\,\Pi_\theta = \Pi_\theta - \theta,$$

Γ_{a^2} and Π_θ are any indep. r.v.'s s.t.

$$\Gamma_{a^2} \sim \mathrm{N}(0, a^2) \text{ and } \Pi_\theta \sim \mathrm{Pois}(\theta).$$

BH and PU exp. bounds: probabilistic interpretation, continued

Remark

So, the BH and PU ineqs. can be viewed as the generalized moment comparison inequalities

$$\mathsf{E}\, f(S) \leqslant \mathsf{E}\, f(y\tilde{\Pi}_{\sigma^2/y^2}) \quad \textit{and}$$

$$\mathsf{E}\, f(S) \leqslant \mathsf{E}\, f\big(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}\big),$$

over the class of all incr. exp. moment functs.
$\mathbb{R} \ni x \mapsto f(x) = e^{\lambda x}$, $\lambda > 0$.

Remark (variance apportionment)

Of the total variance $\sigma^2 = \operatorname{Var}\big(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}\big)$*:*
$(1-\varepsilon)\sigma^2 = \operatorname{Var}\Gamma_{(1-\varepsilon)\sigma^2}$ *and* $\varepsilon\sigma^2 = \operatorname{Var}\big(y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}\big)$, *the variances of the light-tail centered-Gaussian and heavy-tail centered-Poisson components, resp.*

Be: improvement of BH

Bentkus '02, '04 extended the BH ineq.

$$\mathsf{E}\, f(S) \leqslant \mathsf{E}\, f(y\tilde{\Pi}_{\sigma^2/y^2})$$

from incr. exp. f to all f of the form $f(x) \equiv (x-t)_+^2$; hence, to all $f \in \mathcal{H}_+^2$.
So, by our optimal tail comparison theorem, $\forall x \geqslant 0$

$$\mathsf{P}(S \geqslant x) \leqslant \mathrm{Be}(x) := P_2(y\tilde{\Pi}_{\sigma^2/y^2}; x) \leqslant c_{2,0}\ \mathsf{P}^{\mathrm{LC}}(y\tilde{\Pi}_{\sigma^2/y^2} \geqslant x);$$

note: $c_{2,0} = e^2/2 = 3.69\ldots$.
Since $\mathcal{H}_+^2$ contains all incr. exp. functs., $\mathrm{Be}(x)$ improves $\mathrm{BH}(x)$.
Similar results for stochastic integrals: Klein, Ma and Privault '06.

What is done in this paper

The main result is the new bound Pin:

$$\begin{tikzcd}
\text{BH} \arrow[r, "r"] \arrow[d, "i"] & \text{PU} \arrow[d, "i"] \\
\text{Be} \arrow[r, "pr"] & \text{Pin}
\end{tikzcd}$$

where

i :=improvement — by using the larger classes $\mathcal{H}_+^\alpha$ of moment functs. instead of the class of incr. exp. functs.

r :=refinement — by taking truncated 3rd moments into account

pr :=partial refinement — by taking truncated 3rd moments into account, but having to use the somewhat smaller class $\mathcal{H}_+^3$ instead of $\mathcal{H}_+^2$; however, $\mathcal{H}_+^3$ cannot be replaced by the larger class $\mathcal{H}_+^p$ for any $p \in (0,3)$.

Possibly not zero-mean r.v.'s

Now, $X_1, \ldots, X_n$ are indep. but possibly not zero-mean r.v.'s; again, $S := X_1 + \cdots + X_n$.

Main theorem

Theorem (Main)

Take any $\sigma > 0$, $y > 0$, $\beta > 0$ *s.t.*

$$\varepsilon := \frac{\beta}{\sigma^2 y} \in (0,1).$$

Suppose that

$$\sum_i \operatorname{E} X_i^2 \leqslant \sigma^2, \quad \sum_i \operatorname{E}(X_i)_+^3 \leqslant \beta, \quad \operatorname{E} X_i \leqslant 0, \quad \textit{and } X_i \leqslant y$$

a.s. $\forall i$. *Then*

$$\operatorname{E} f(S) \leqslant \operatorname{E} f\left(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}\right) \quad \forall f \in \mathcal{H}_+^3.$$

Exactness properties

Proposition (Exactness for each f)

For each triple (σ, y, β) *as in Theorem (Main) and each* $f \in \mathcal{H}^3_+$, *the upper bound* $\mathsf{E}\, f\big(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}\big)$ *on* $\mathsf{E}\, f(S)$ *is exact.*

Proposition (Exactness in p)

For any given $p \in (0, 3)$, *one cannot replace* $\mathcal{H}^3_+$ *in Theorem (Main) by the larger class* $\mathcal{H}^p_+$.

Corollary: the upper bound Pin(x) on the tail

From Theorem (Main) and the optimal comparison remark, one immediately obtains

Corollary (Upper bound on the tail)

Under the conditions of Theorem (Main), $\forall x \in \mathbb{R}$

$$\begin{aligned} \mathsf{P}(S \geqslant x) \leqslant \mathrm{Pin}(x) &:= P_3(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}; x) \\ &\leqslant c_{3,0}\ \mathsf{P}^{\mathrm{LC}}(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2} \geqslant x); \end{aligned}$$

note: $c_{3,0} = 2e^3/9 = 4.46\ldots$.

Sketch of proof

Remark

Since the class $\mathcal{H}^3_+$ of generalized moment functs. is shift-invariant, it is enough to prove Theorem (Main) just for $n = 1$.

Fix any $\sigma > 0$ and $y > 0$.

For any $a \geqslant 0$ and $b > 0$, let $X_{a,b}$ denote any r.v. with the unique zero-mean distr. on the two-point set $\{-a, b\}$.

Lemma: possible values of $\mathsf{E}\, X_+^3$

Lemma (Possible values of $\mathsf{E}\, X_+^3$)

(i) *For any r.v.* X *s.t.* $X \leqslant y$ *a.s.,* $\mathsf{E}\, X \leqslant 0$, *and* $\mathsf{E}\, X^2 \leqslant \sigma^2$,

$$\mathsf{E}\, X_+^3 \leqslant \frac{y^3\sigma^2}{y^2+\sigma^2}.$$

(ii) *For any*

$$\beta \in \Big(0, \frac{y^3\sigma^2}{y^2+\sigma^2}\Big]$$

$\exists!(a,b) \in (0,\infty)\times(0,\infty)$ *s.t.* $X_{a,b} \leqslant y$ *a.s.,* $\mathsf{E}\, X_{a,b}^2 = \sigma^2$, *and* $\mathsf{E}(X_{a,b})_+^3 = \beta$.

In particular, the ineq. in part (i) is exact.

2-point zero-mean distrs. are extremal

Lemma (2-point zero-mean distrs. are extremal)

Fix any $w \in \mathbb{R}$, $y > 0$, $\sigma > 0$, *and* β *s.t.* $\beta \in \left(0, \frac{y^3\sigma^2}{y^2+\sigma^2}\right]$, *and let* (a, b) *be the unique pair as in the previous lemma. Then*

$$\max\{\mathsf{E}(X-w)_+^3 \colon X \leqslant y \text{ a.s.}, \mathsf{E}\,X \leqslant 0, \mathsf{E}\,X^2 \leqslant \sigma^2, \mathsf{E}\,X_+^3 \leqslant \beta\}$$

$$= \begin{cases} \mathsf{E}(X_{a,b}-w)_+^3 & \text{if } w \leqslant 0, \\ \mathsf{E}(X_{\tilde a,\tilde b}-w)_+^3 & \text{if } w \geqslant 0, \end{cases}$$

where $\tilde b := y$ *and* $\tilde a := \frac{\beta y}{y^3-\beta}$. *At that,* $\tilde a > 0$, $X_{\tilde a,\tilde b} \leqslant y$ *a.s.,* $\mathsf{E}\,X_{\tilde a,\tilde b} = 0$, *and* $\mathsf{E}(X_{\tilde a,\tilde b})_+^3 = \beta$, *but one can only say that* $\mathsf{E}\,X_{\tilde a,\tilde b}^2 \leqslant \sigma^2$, *and the latter inequality is strict if* $\beta \neq \frac{y^3\sigma^2}{y^2+\sigma^2}$.

Monotonicity in σ and β

Lemma (Monotonicity in σ and β)

Take any $\sigma_0, \beta_0, \sigma, \beta$ *s.t.*
$0 \leqslant \sigma_0 \leqslant \sigma$, $0 \leqslant \beta_0 \leqslant \beta$,
$\beta_0 \leqslant \sigma_0^2 y$, *and* $\beta \leqslant \sigma^2 y$. *Then*

$$\mathsf{E}\, f(\Gamma_{\sigma_0^2-\beta_0/y} + y\,\tilde{\Pi}_{\beta_0/y^3}) \leqslant \mathsf{E}\, f(\Gamma_{\sigma^2-\beta/y} + y\,\tilde{\Pi}_{\beta/y^3}) \qquad (1)$$

$\forall f \in \mathcal{H}_+^2$, *and hence* $\forall f \in \mathcal{H}_+^3$.

Main lemma

Lemma (Main)

Let X *be any r.v such that* $X \leqslant y$ *a.s.,* $\operatorname{E} X \leqslant 0$, $\operatorname{E} X^2 \leqslant \sigma^2$, *and* $\operatorname{E} X_+^3 \leqslant \beta$, *where* $\beta \in \left(0, \frac{y^3\sigma^2}{y^2+\sigma^2}\right]$. *Then*

$$\operatorname{E} f(X) \leqslant \operatorname{E} f(\Gamma_{\sigma^2-\beta/y} + y\,\tilde{\Pi}_{\beta/y^3}) \quad \forall f \in \mathcal{H}_+^3.$$

Sketch of proof By the "2-point zero-mean distrs. are extremal" lemma and the "monotonicity in σ and β" lemma, w.l.o.g. $X = X_{a_0,b_0}$ for some $a_0 > 0$ and $b_0 > 0$. Also, w.l.o.g. $f(x) \equiv (x-w)_+^3$. Also, by rescaling, w.l.o.g. $y = 1$.

Main idea of the proof of Lemma (Main): *infinitesimal spin-off*

The initial infinitesimal step:

Start with the r.v. X_{a_0,b_0}. Decrease a_0 and b_0 simultaneously by infinitesimal amounts $\Delta a > 0$ and $\Delta b > 0$ so that $\mathsf{E}(X_{a_0,b_0} - w)_+^3 \leqslant \mathsf{E}(X_{a,b} + X_{\Delta_1,\Delta_1} + X_{\Delta_2,1} - w)_+^3 \ \forall w \in \mathbb{R}$, where $X_{a,b}, X_{\Delta_1,\Delta_1}, X_{\Delta_2,1}$ are indep., $a = a_0 - \Delta a$ and $b = b_0 - \Delta b$, and $0 < \Delta_1 \approx 0$ and $0 < \Delta_2 \approx 0$ are chosen, together with Δa and Δb, so that to keep the balance of the total variance and that of the positive-part third moments closely enough:
$\mathsf{E}\, X_{a,b}^2 + \mathsf{E}\, X_{\Delta_1,\Delta_1}^2 + \mathsf{E}\, X_{\Delta_2,1}^2 \approx \mathsf{E}\, X_{a_0,b_0}^2$
and $\mathsf{E}(X_{a,b})_+^3 + \mathsf{E}(X_{\Delta_1,\Delta_1})_+^3 + \mathsf{E}(X_{\Delta_2,1})_+^3 \approx \mathsf{E}(X_{a_0,b_0})_+^3$.
Refer to X_{Δ_1,Δ_1} and $X_{\Delta_2,1}$ as the symm. and highly asymm. infinitesimal spin-offs, resp.

Infinitesimal spin-off: continued

Continue decreasing a and b while "spinning off" the indep. pairs of indep. infinitesimal spin-offs X_{Δ_1,Δ_1} and $X_{\Delta_2,1}$, at that keeping the balance of the total variance and that of the positive-part third moments, as described. Stop when $X_{a,b}=0$ a.s., i.e., when a or b is decreased to 0 (if ever); such a termination point is indeed attainable. Then the sum of all the symm. indep. infinitesimal spin-offs X_{Δ_1,Δ_1} will have a centered Gaussian distr., while the sum of the highly asymmetric spin-offs $X_{\Delta_2,1}$'s will give a centered Poisson component. At that, the balances of the variances and positive-part third moments will each be kept (the infinitesimal X_{Δ_1,Δ_1}'s will provide in the limit a total zero contribution to the balance of the positive-part third moments).

Formalizing the spin-off idea, with a time-changed Lévy process

Introduce a family of r.v.'s of the form

$$\eta_b := X_{a(b),b} + \xi_{\tau(b)} \quad \text{for} \quad b \in [\varepsilon, b_0], \qquad \text{where}$$

$$\varepsilon := \beta/\sigma^2 = b_0^2/(b_0 + a_0) < b_0,$$

$$a(b) := (b/\varepsilon - 1)b, \quad \tau(b) := a_0 b_0 - a(b)b, \qquad (\text{balances})$$

$$\xi_t := W_{(1-\varepsilon)t} + \tilde{\Pi}_{\varepsilon t},$$

$W.$ and $\tilde{\Pi}.$ are indep. standard Wiener and centered standard Poisson processes, indep. of $X_{a(b),b}$ for each $b \in [\varepsilon, b_0]$. Note: $a(b_0) = a_0$ and $a(\varepsilon) = 0$, $\tau(b_0) = 0$ and $\tau(\varepsilon) = a_0 b_0 = \sigma^2$, so that

$$\eta_{b_0} = X_{a_0,b_0} \quad \text{and} \quad \eta_\varepsilon = W_{(1-\varepsilon)\sigma^2} + \tilde{\Pi}_{\varepsilon\sigma^2}.$$

Thus, it's enough to show that $\mathsf{E}(\eta_b - w)_+^3$ decr. in $b \in [\varepsilon, b_0]$, for each $w \in \mathbb{R}$.

PU computation

Proposition (PU(x) computation)

For all $\sigma > 0$, $y > 0$, $\varepsilon \in (0, 1)$, *and* $x \geqslant 0$

$$\begin{aligned}\mathrm{PU}(x) &= e^{-\lambda_x x}\,\mathrm{PU}_{\exp}(\lambda_x)\\ &= \exp\frac{(1-\varepsilon)^2(w_x+1)^2 - (\varepsilon + xy/\sigma^2)^2 - (1-\varepsilon^2)}{2(1-\varepsilon)y^2/\sigma^2},\end{aligned}$$

$$\lambda_x := \frac{1}{y}\Big(\frac{\varepsilon + xy/\sigma^2}{1-\varepsilon} - w_x\Big), \quad w_x := L\Big(\frac{\varepsilon}{1-\varepsilon}\exp\frac{\varepsilon + xy/\sigma^2}{1-\varepsilon}\Big),$$

and L *is the Lambert product-log funct.:* $\forall z \geqslant 0$, $w = L(z)$ *is the only real root of the equation* $we^w = z$.
Moreover, λ_x *incr. in* x *from* 0 *to* ∞ *as* x *does so.*

So, PU(x) is about as easy to compute as BH(x).

Computation of $\mathrm{Be}(x)$ and $\mathrm{Pin}(x)$

Recall:
$\mathrm{Be}(x) := P_2(y\tilde{\Pi}_{\sigma^2/y^2}; x)$ and
$\mathrm{Pin}(x) := P_3(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}; x)$, where

$$P_\alpha(\eta; x) := \inf_{t\in(-\infty,x)} \frac{\mathsf{E}(\eta-t)_+^\alpha}{(x-t)^\alpha}.$$

An efficient procedure to compute $P_\alpha(\eta; x)$ in general was given in Pinelis '98.
In the case of $\mathrm{Be}(x) = P_2(y\tilde{\Pi}_{\sigma^2/y^2}; x)$, this general procedure can be much simplified. Indeed, if α is natural and $\cdots < d_k < d_{k+1} < \cdots$ are the atoms of the distr. of η, then $\mathsf{E}(\eta-t)_+^\alpha$ can be easily expressed for $t \in [d_k, d_{k+1})$ in terms of the truncated moments $\mathsf{E}(\eta-d_k)_+^j$ with $j = 0, \ldots, \alpha$.

Computation of Pin(x)

For $\mathrm{Pin}(x) = P_3(\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}; x)$, there is no such nice localization property as for $\mathrm{Be}(x) = P_2(y\tilde{\Pi}_{\sigma^2/y^2}; x)$, since the distr. of the r.v. $\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}$ is not discrete.

A good way to compute $\mathrm{Pin}(x)$ turns out to be to express the positive-part moments $\mathsf{E}(\eta - t)_+^\alpha$ for $\eta = \Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}$ in terms of the Fourier or Fourier-Laplace transform of the distribution of η. Such expressions were developed in Pinelis '09 (with this specific motivation in mind). A reason for this approach to work is that the Fourier-Laplace transform of the distribution of the r.v. $\Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}$ has a simple expression.

Expressions for the positive-part moments in terms of the Fourier or Fourier-Laplace transform

$$\mathsf{E}\,X_+^p = \frac{\Gamma(p+1)}{\pi}\int_0^\infty \mathfrak{Re}\,\frac{\mathsf{E}\,e_j\big((s+it)X\big)}{(s+it)^{p+1}}\,\mathrm{d}t,$$

where $p \in (0,\infty)$, $s \in (0,\infty)$, Γ is the Gamma function, $\mathfrak{Re}\,z :=$ the real part of z, $i = \sqrt{-1}$, $j = -1, 0, \dots, \ell$, $\ell := \lceil p-1 \rceil$, $e_j(u) := e^u - \sum_{m=0}^{j} \frac{u^m}{m!}$, and X is any r.v. s.t. $\mathsf{E}\,|X|^{j_+} < \infty$ and $\mathsf{E}\,e^{sX} < \infty$.
Also,

$$\mathsf{E}\,X_+^p = \frac{\mathsf{E}\,X^k}{2}\,\mathrm{I}\{p \in \mathbb{N}\} + \frac{\Gamma(p+1)}{\pi}\int_0^\infty \mathfrak{Re}\,\frac{\mathsf{E}\,e_\ell(itX)}{(it)^{p+1}}\,\mathrm{d}t,$$

where $k := \lfloor p \rfloor$ and X is any r.v. such that $\mathsf{E}\,|X|^p < \infty$. Of course, these formulas are to be applied here to $X = \Gamma_{(1-\varepsilon)\sigma^2} + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2} - w$, $w \in \mathbb{R}$.

Comparison

Compare the bounds BH, PU, Be, and Pin, and also the Cantelli bound

$$\mathrm{Ca}(x) := \mathrm{Ca}_{\sigma^2}(x) := \frac{\sigma^2}{\sigma^2 + x^2}$$

and the best exp. bound

$$\mathrm{EN}(x) := \mathrm{EN}_{\sigma^2}(x) \inf_{\lambda>0} e^{-\lambda x}\,\mathsf{E}\,e^{\lambda\Gamma_{\sigma^2}} = \exp\Big\{-\frac{x^2}{2\sigma^2}\Big\}$$

on the tail of $\mathrm{N}(0,\sigma^2)$; of course, in general $\mathrm{EN}(x)$ is *not* an upper bound on $\mathsf{P}(S \geqslant x)$.

The bound $\mathrm{Ca}(x)$ is optimal in its own terms.

Proposition

Take any $x \in [0,\infty)$, $\sigma \in (0,\infty)$, *and r.v.'s* ξ *and* η *s.t.* $\mathsf{E}\,\xi \leqslant 0 = \mathsf{E}\,\eta$ *and* $\mathsf{E}\,\xi^2 \leqslant \mathsf{E}\,\eta^2 = \sigma^2$. *Then*

$$\mathsf{P}(\xi \geqslant x) \leqslant \mathrm{Ca}(x) = \inf_{t\in(-\infty,x)} \frac{\mathsf{E}(\eta-t)^2}{(x-t)^2}.$$

Comparison: ineqs.

Proposition

For all $x > 0$, $\sigma > 0$, $y > 0$, *and* $\varepsilon \in (0, 1)$,

- (I) $\mathrm{Pin}(x) \leqslant \mathrm{PU}(x) \leqslant \mathrm{BH}(x)$ *and* $\mathrm{Be}(x) \leqslant \mathrm{Ca}(x) \wedge \mathrm{BH}(x)$;
- (II) $\mathrm{Be}(x) = \mathrm{Ca}(x)$ *for all* $x \in [0, y]$;
- (III) $\mathrm{BH}(x)$ *increases from* $\mathrm{EN}(x)$ *to* 1 *as* y *increases from* 0 *to* ∞;
- (IV) $\exists u_{y/\sigma} \in (0, \infty)$ *s.t.* $\mathrm{Ca}(x) < \mathrm{BH}(x)$ *if* $x \in (0, \sigma u_{y/\sigma})$ *and* $\mathrm{Ca}(x) > \mathrm{BH}(x)$ *if* $x \in (\sigma u_{y/\sigma}, \infty)$; *moreover,* $u_{y/\sigma}$ *incr. from* $u_{0+} = 1.585\dots$ *to* ∞ *as* y/σ *incr. from* 0 *to* ∞; *in particular,* $\mathrm{Ca}(x) < \mathrm{EN}(x)$ *if* $x/\sigma \in (0, 1.585)$ *and* $\mathrm{Ca}(x) > \mathrm{EN}(x)$ *for* $x/\sigma \in (1.586, \infty)$.
- (V) $\mathrm{PU}(x)$ *incr. from* $\mathrm{EN}(x)$ *to* $\mathrm{BH}(x)$ *as* ε *incr. from* 0 *to* 1.

Comparison: identity

Proposition

For all $\sigma > 0$, $y > 0$, $\varepsilon \in (0,1)$, *and* $x > 0$

$$\begin{aligned}\mathsf{PU}(x) &= \max_{\alpha\in(0,1)} \mathsf{EN}_{(1-\varepsilon)\sigma^2}((1-\alpha)x)\,\mathsf{BH}_{\varepsilon\sigma^2,y}(\alpha x)\\ &= \mathsf{EN}_{(1-\varepsilon)\sigma^2}((1-\alpha_x)x)\,\mathsf{BH}_{\varepsilon\sigma^2,y}(\alpha_x x),\end{aligned}$$

where α_x *is the only root in* $(0,1)$ *of the equation*
$\frac{(1-\alpha)x^2}{(1-\varepsilon)\sigma^2} - \frac{x}{y}\ln\left(1+\frac{\alpha x y}{\varepsilon\sigma^2}\right) = 0.$
Moreover, α_x *incr. from* ε *to* 1 *as* x *incr. from* 0 *to* ∞.

So, the bound $\mathsf{PU}(x)$ is the product of the best exp. upper bounds on the tails $\mathsf{P}\left(\Gamma_{(1-\varepsilon)\sigma^2} \geqslant (1-\alpha)x\right)$ and $\mathsf{P}\left(\tilde{\Pi}_{\varepsilon\sigma^2} \geqslant \alpha x\right)$ — for some $\alpha \in (0,1)$ $\big($in fact, the $\alpha \in (\varepsilon,1)\big)$.
This proposition is useful in establishing asymptotics of $\mathsf{PU}(x)$.

Comparison: asymptotics for large $x > 0$

Proposition

For any fixed $\sigma > 0$, $y > 0$, *and* $\varepsilon \in (0, 1)$, *and all* $x \geqslant 0$

$$\mathrm{Pin}(x) \leqslant \mathrm{PU}(x) = (\varepsilon + o(1))^{x/y}\,\mathrm{Be}(x) \leqslant (\varepsilon + o(1))^{x/y}\,\mathrm{BH}(x)$$

as $x \to \infty$.

That is, for large x, the bound $\mathrm{PU}(x)$ and, hence, the better bound $\mathrm{Pin}(x)$ are each exponentially better than $\mathrm{Be}(x)$ and hence than $BH(x)$ — especially when $\varepsilon << 1$.

Graphic comparison for moderate deviations: $x \in [0,3]$ or $x \in [0,4]$

Here, σ is normalized to be 1. In the next 4 frames, the graphs $G(P) := \{\big(x, \log_{10} \frac{P(x)}{\mathrm{BH}(x)}\big) : 0 < x \leqslant x_{\max}\}$ for $P = \mathrm{Ca}, \mathrm{PU}, \mathrm{Be}, \mathrm{Pin}$, with the benchmark BH, will be shown, for $\varepsilon \in \{0.1, 0.9\}$, $y \in \{0.1, 1\}$, and $x_{\max} = 3$ or 4, depending on whether $y = 0.1$ (little skewed-to-the-right X_i's) or $y = 1$ (much skewed-to-the-right X_i's).

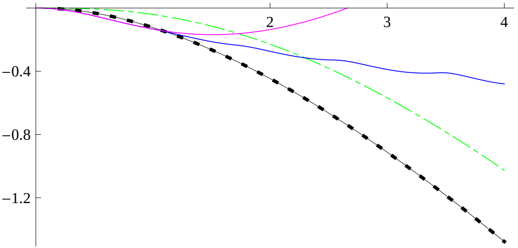
¶ for such choices of $x_{\max}$, the values of $\mathrm{BH}(x_{\max}) \approx 0.016$ or 0.017, whether $y = 0.1$ or $y = 1$.

¶ $G(\mathrm{Ca})$ is shown only on the interval $(0, u_y)$, on which $\mathrm{Ca} < \mathrm{BH}$, i.e., $\log_{10} \frac{\mathrm{Ca}}{\mathrm{BH}} < 0$.

¶ for $y = 1$, $\mathrm{Ca}(x) < \mathrm{BH}(x)$ for all $x \in (0, 2.66)$.

¶ For Pin, actually two approx. graphs are shown: the dashed and thin solid lines – produced using the Fourier-Laplace and Fourier formulas.

Comparison: $x \in [0,4]$, $\varepsilon = 0.1$, $y = 1$



BH ~ 0

(BH(4) ≈ 0.017)

Ca

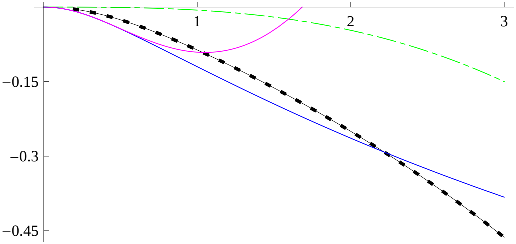
Be

PU

Pin

If the weight of the Poisson component is small ($\varepsilon = 0.1$) and the Poisson component is quite distinct from the Gaussian component ($y = 1$), then Be(x) is about 9.93 times worse (i.e., greater) than Pin(x) at $x = 4$. Moreover, for these values of ε and y, even the bound PU(x) is better than Be(x) already at about $x = 2.5$.

Comparison: $x \in [0,3]$, $\varepsilon = 0.1$, $y = 0.1$



BH ~ 0

(BH(4) ≈ 0.016)

Ca

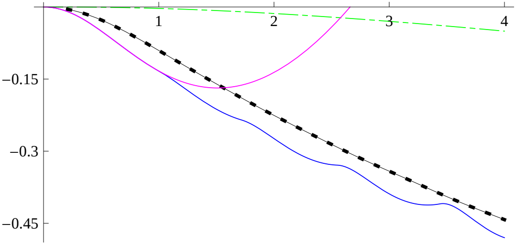
Be

PU

Pin

If the weight of the Poisson component is small ($\varepsilon = 0.1$) and the Poisson component is close to the Gaussian component ($y = 0.1$), then Be(x) is still about 20% greater than Pin(x) at $x = 3$.

Comparison: $x \in [0, 4]$, $\varepsilon = 0.9$, $y = 1$



BH ~ 0

(BH(4) ≈ 0.017)

Ca

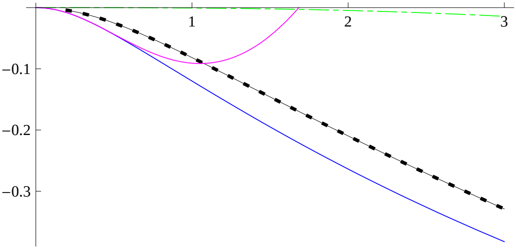
Be

PU

Pin

If the weight of the Poisson component is large ($\varepsilon = 0.9$) and the Poisson component is quite distinct from the Gaussian component ($y = 1$), then Be(x) is about 8% better than Pin(x) at $x = 4$. For $x \in [0, 4]$, Pin(x) and Be(x) are close to each other and both are significantly better than either BH(x) or PU(x) (which latter are also close to each other).

Comparison: $x \in [0, 4]$, $\varepsilon = 0.9$, $y = 0.1$



BH ∼ 0

(BH(4) ≈ 0.016)

Ca

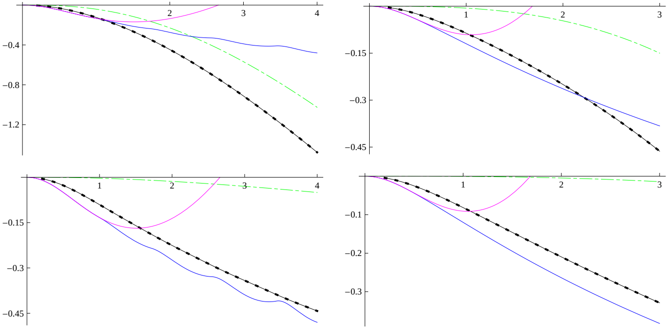
Be

PU

Pin

If the weight of the Poisson component is large ($\varepsilon = 0.9$) and the Poisson component is close to the Gaussian component ($y = 0.1$), then Be(x) is about 12% better than Pin(x) at $x = 3$. For $x \in [0, 3]$, Pin(x) and Be(x) are close to each other and both are significantly better than either BH(x) or PU(x) (which latter are *very* close to each other).

Comparison: Graphics grid



Row 1: $\varepsilon = 0.1$: heavy-tail Poisson component of little weight
Row 2: $\varepsilon = 0.9$: heavy-tail Poisson component of large weight
Column 1: $y = 1$: distrs. of the X_i's may be much skewed to the right
Column 2: $y = 0.1$: distrs. of the X_i's may be only a little skewed to the right.

Thank you!