On the Bennett-Hoeffding inequality

Iosif Pinelis $^{1,2,3}$

1Department of Mathematical Sciences
Michigan Technological University
2Supported by NSF grant DMS-0805946

October 14, 2009
1 Introduction

2 Main results

3 Sketch of proof

4 Computation of bounds

5 Comparison of bounds
$X_1, \ldots, X_n$: indep. r.v.'s s.t. $X_i \leq y$ a.s. for some $y > 0$;
$E X_i \leq 0$;
$\sum_i E X_i^2 \leq \sigma^2$.
$S := X_1 + \cdots + X_n$. 
Classes of (generalized moment) functions $f : \mathbb{R} \to \mathbb{R}$

$f \in \mathcal{E} \iff \exists \lambda > 0 \ \forall u \in \mathbb{R} \ f(u) = e^{\lambda u};$

$f \in \mathcal{H}_+^\alpha \iff \exists \mu \geq 0 \ \forall u \in \mathbb{R} \ f(u) = \int_{-\infty}^{\infty} (u - t)^\alpha_+ \mu(dt).$

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

For $\alpha = 1, 2, \ldots$; $f \in \mathcal{H}_+^\alpha$ iff $f^{(\alpha-1)}$ is convex and $f^{(j)}(-\infty) = 0$ for $j = 0, \ldots, \alpha - 1.$

$$\bigcap_{\alpha > 0} \mathcal{H}_+^\alpha = \{ f : f(x) = \int_{(0,\infty)} e^{tx} \mu(dt) \ \forall x \in \mathbb{R} \}.$$
Let $\Gamma_{a^2}$ and $\Pi_\theta$ be any indep. r.v.’s s.t.

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

$$\tilde{\Pi}_\theta := \Pi_\theta - E\Pi_\theta = \Pi_\theta - \theta.$$
Main theorem

**Theorem (Main)**

Take any $\beta > 0$ s.t.

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

Suppose that

$$\sum_i \mathbb{E}(X_i)^3_+ \leq \beta.$$

Then

$$\mathbb{E} f(S) \leq \mathbb{E} f\left(\Gamma_{(1-\varepsilon)}\sigma^2 + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}\right) \quad \forall f \in \mathcal{H}_+^3.$$
Proposition (Exactness for each $f$)

For each triple $(\sigma, y, \beta)$ as in Theorem (Main) and each $f \in \mathcal{H}_+^3$, the upper bound $E f \left( \Gamma_{(1-\varepsilon)} \sigma^2 + y \tilde{\Pi}_{\varepsilon} \sigma^2 / y^2 \right)$ on $E f(S)$ 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 $\mathcal{H}_+^p$. 
Related preceding results – all of the form:

$$\forall f \in \mathcal{F} \quad \sup E f(S) = E f(\eta),$$

sup over all indep. $X_i$’s as before, with the cond. $\sum E(X_i)_+^3 \leq \beta$ imposed or not, and where the class $\mathcal{F}$ of functions and the r.v. $\eta$ are as in the following table:

<table>
<thead>
<tr>
<th>Bound</th>
<th>$\mathcal{F}$</th>
<th>$\sum E(X_i)_+^3 \leq \beta$ imposed?</th>
<th>$\eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BH</td>
<td>$\mathcal{E}$</td>
<td>no</td>
<td>$y\tilde{\Pi}_{\sigma^2/y^2}$</td>
</tr>
<tr>
<td>PU</td>
<td>$\mathcal{E}$</td>
<td>yes</td>
<td>$\Gamma(1-\varepsilon)\sigma^2 + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}$</td>
</tr>
<tr>
<td>Be</td>
<td>$\mathcal{H}_+^2$</td>
<td>no</td>
<td>$y\tilde{\Pi}_{\sigma^2/y^2}$</td>
</tr>
<tr>
<td>Pin</td>
<td>$\mathcal{H}_+^3$</td>
<td>yes</td>
<td>$\Gamma(1-\varepsilon)\sigma^2 + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2}$</td>
</tr>
</tbody>
</table>
Corollary (Upper bound on the tail)

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

\[
P(S \geq x) \leq \text{Pin}(x) := P_{\mathcal{H}_+^3}^{\text{c}} \left( \Gamma (1-\varepsilon) \sigma^2 + y \tilde{\Pi}_{\varepsilon} \sigma^2 / y^2 ; x \right) \\
\leq c_{3,0} \ \text{P}^{\text{LC}} \left( \Gamma (1-\varepsilon) \sigma^2 + y \tilde{\Pi}_{\varepsilon} \sigma^2 / y^2 \geq x \right),
\]

\[c_{3,0} = 2e^3 / 9 = 4.46 \ldots .\]

Here, \( \text{P}^{\text{LC}}(\eta \geq \cdot) \) is the least log-concave majorant of \( \text{P}(\eta \geq \cdot) \), and

\[P_{\mathcal{F}}(\eta; x) = \inf \left\{ \frac{\mathbb{E} f(\eta)}{f(x)} : f \in \mathcal{F}, \ f(x) > 0 \right\}, \]

the best upper bound on \( \text{P}(S \geq x) \) based on comparison \( \mathbb{E} f(S) \leq \mathbb{E} f(\eta) \) for all \( f \in \mathcal{F} \).
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 $a \geq 0$ 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 $E X^3_+$)

(i) For any r.v. $X$ s.t. $X \leq y$ a.s., $E X \leq 0$, and $E X^2 \leq \sigma^2$,

$$E X^3_+ \leq \frac{y^3 \sigma^2}{y^2 + \sigma^2}.$$

(ii) For any

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

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

In particular, the ineq. in part (i) is exact.
Lemma (2-point zero-mean distr. are extremal)

Fix any \( w \in \mathbb{R}, y > 0, \sigma > 0, \) and \( \beta \) 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

\[
\max \{ E(X - w)^3_+: X \leq y \text{ a.s., } E X \leq 0, E X^2 \leq \sigma^2, E X^3_+ \leq \beta \} = \begin{cases} 
E(X_{a,b} - w)^3_+ & \text{if } w \leq 0, \\
E(X_{\tilde{a},\tilde{b}} - w)^3_+ & \text{if } w \geq 0,
\end{cases}
\]

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

Take any $\sigma_0, \beta_0, \sigma, \beta$ s.t.

\[ 0 \leq \sigma_0 \leq \sigma, \ 0 \leq \beta_0 \leq \beta, \]

\[ \beta_0 \leq \sigma_0^2 y, \text{ and } \beta \leq \sigma^2 y. \]

Then

\[ \mathbb{E} f(\Gamma_{\sigma_0^2-\beta_0/y} + y \tilde{\Pi}_{\beta_0/y^3}) \leq \mathbb{E} f(\Gamma_{\sigma^2-\beta/y} + y \tilde{\Pi}_{\beta/y^3}) \] (1)

\[ \forall f \in \mathcal{H}_+^2, \text{ and hence } \forall f \in \mathcal{H}_+^3. \]
**Main lemma**

**Lemma (Main)**

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

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

**Sketch of proof**  
By the “2-point zero-mean distr. are extremal” lemma and the “monotonicity in $\sigma$ and $\beta$” 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. $y = 1$. 
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
$$E(X_{a_0,b_0} - w)_+^3 \leq E(X_a,b + X_{\Delta_1,\Delta_1} + X_{\Delta_2,1} - w)_+^3 \quad \forall w \in \mathbb{R},$$
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 $\Delta a$ and $\Delta b$, so that to keep the balance of the total variance and that of the positive-part third moments closely enough:
$$E X_{a,b}^2 + E X_{\Delta_1,\Delta_1}^2 + E X_{\Delta_2,1}^2 \approx E X_{a_0,b_0}^2$$
and
$$E(X_{a,b})_+^3 + E(X_{\Delta_1,\Delta_1})_+^3 + E(X_{\Delta_2,1})_+^3 \approx E(X_{a_0,b_0})_+^3.$$
Refer to $X_{\Delta_1,\Delta_1}$ and $X_{\Delta_2,1}$ as the symm. and highly asymm. infinitesimal spin-offs, resp.
Continue decreasing $a$ and $b$ while “spinning off” the indep. pairs of indep. infinitesimal spin-offs $X_{\Delta_1, \Delta_1}$ 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 $X_{\Delta_1, \Delta_1}$ 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 $X_{\Delta_1, \Delta_1}$’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
\[ \eta_b := X_{a(b),b} + \xi_{\tau(b)} \quad \text{for} \quad b \in [\varepsilon, b_0], \quad \text{where} \]
\[ \varepsilon := \beta / \sigma^2 = b_0^2 / (b_0 + a_0) < b_0, \]
\[ a(b) := (b / \varepsilon - 1)b, \quad \tau(b) := a_0 b_0 - a(b)b, \quad \text{(balances)} \]
\[ \xi_t := \mathcal{W}_{(1-\varepsilon)}t + \tilde{\Pi}_{\varepsilon t}, \]
\( \mathcal{W} \) 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
\[ \eta_{b_0} = X_{a_0,b_0} \quad \text{and} \quad \eta_{\varepsilon} = \mathcal{W}_{(1-\varepsilon)} \sigma^2 + \tilde{\Pi}_{\varepsilon \sigma^2}. \]
Thus, it’s enough to show that \( E(\eta_b - w)^3 \) decr. in \( b \in [\varepsilon, b_0] \), for each \( w \in \mathbb{R} \).
Proposition (PU($x$) computation)

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

$$PU(x) = e^{-\lambda_x x} PU_{\exp}(\lambda_x)$$
$$= \exp \left( \frac{(1 - \varepsilon)^2 (w_x + 1)^2 - (\varepsilon + xy/\sigma^2)^2 - (1 - \varepsilon)^2}{2(1 - \varepsilon)y^2/\sigma^2} \right),$$

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

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

So, PU($x$) is about as easy to compute as BH($x$).
Recall: \( \text{Be}(x) := P_2(y\tilde{\Pi}_{\sigma^2/y^2};x) \) and 
\( \text{Pin}(x) := P_3(\Gamma(1-\varepsilon)\sigma^2 + y\tilde{\Pi}_{\varepsilon\sigma^2/y^2};x) \), where

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

An efficient procedure to compute \( P_{\alpha}(\eta; x) \) in general was given in Pinelis '98.

In the case of \( \text{Be}(x) = P_2(y\tilde{\Pi}_{\sigma^2/y^2};x) \), this general procedure can be much simplified. Indeed, if \( \alpha \) is natural and 
\( \cdots < d_k < d_{k+1} < \cdots \) are the atoms of the distr. of \( \eta \), then
\( E(\eta - t)^{\alpha+} \) can be easily expressed for \( t \in [d_k, d_{k+1}) \) in terms of the truncated moments \( E(\eta - d_k)^{j+} \) with \( j = 0, \ldots, \alpha \).
For $\text{Pin}(x) = P_3(\Gamma_{(1-\varepsilon)\sigma^2 + y\tilde{\Pi}_{\varepsilon}\sigma^2/y^2}; x)$, there is no such nice localization property as for $\text{Be}(x) = P_2(y\tilde{\Pi}_{\varepsilon}\sigma^2/y^2; x)$, 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 $\text{Pin}(x)$ turns out to be to express the positive-part moments $E(\eta - t)^\alpha_+$ 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 $\eta$. 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

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

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

Also,

\[ E X_+^p = \frac{E X_k^2}{2} \mathbf{1}\{p \in \mathbb{N}\} + \frac{\Gamma(p + 1)}{\pi} \int_0^\infty \Re \frac{E e^{\ell(it)X}}{(it)^{p+1}} \, dt, \]

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

Compare the bounds BH, PU, Be, and Pin, and also the Cantelli bound
\[ \text{Ca}(x) := \text{Ca}_{\sigma^2}(x) := \frac{\sigma^2}{\sigma^2 + x^2} \]
and the best exp. bound
\[ \text{EN}(x) := \text{EN}_{\sigma^2}(x) \inf_{\lambda > 0} e^{-\lambda x} E e^{\lambda \Gamma_{\sigma^2}} = \exp \left\{ -\frac{x^2}{2\sigma^2} \right\} \]
on the tail of \( N(0, \sigma^2) \); of course, in general \( \text{EN}(x) \) is not an upper bound on \( P(S \geq x) \).
The bound \( \text{Ca}(x) \) is optimal in its own terms.

Proposition

Take any \( x \in [0, \infty) \), \( \sigma \in (0, \infty) \), and r.v.'s \( \xi \) and \( \eta \) s.t.
\( E \xi \leq 0 = E \eta \) and \( E \xi^2 \leq E \eta^2 = \sigma^2 \). Then
\[ P(\xi \geq x) \leq \text{Ca}(x) = \inf_{t \in (-\infty, x)} \frac{E(\eta - t)^2}{(x - t)^2}. \]
Proposition

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

(I) $\text{Pin}(x) \leq \text{PU}(x) \leq \text{BH}(x)$ and $\text{Be}(x) \leq \text{Ca}(x) \land \text{BH}(x)$;

(II) $\text{Be}(x) = \text{Ca}(x)$ for all $x \in [0, y]$;

(III) $\text{BH}(x)$ increases from $\text{EN}(x)$ to 1 as $y$ increases from 0 to $\infty$;

(IV) $\exists u_{y/\sigma} \in (0, \infty)$ s.t. $\text{Ca}(x) < \text{BH}(x)$ if $x \in (0, \sigma u_{y/\sigma})$ and $\text{Ca}(x) > \text{BH}(x)$ if $x \in (\sigma u_{y/\sigma}, \infty)$; moreover, $u_{y/\sigma}$ incr. from $u_{0+} = 1.585 \ldots$ to $\infty$ as $y/\sigma$ incr. from 0 to $\infty$; in particular, $\text{Ca}(x) < \text{EN}(x)$ if $x/\sigma \in (0, 1.585)$ and $\text{Ca}(x) > \text{EN}(x)$ for $x/\sigma \in (1.586, \infty)$.

(V) $\text{PU}(x)$ incr. from $\text{EN}(x)$ to $\text{BH}(x)$ as $\varepsilon$ incr. from 0 to 1.
Proposition

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

$$PU(x) = \max_{\alpha \in (0, 1)} EN_{(1-\varepsilon)\sigma^2}((1 - \alpha)x) BH_{\varepsilon \sigma^2, y}(\alpha x)$$

$$= EN_{(1-\varepsilon)\sigma^2}((1 - \alpha_x)x) BH_{\varepsilon \sigma^2, y}(\alpha_x x),$$

where $\alpha_x$ is the only root in $(0, 1)$ of the equation

$$\frac{(1-\alpha)x^2}{(1-\varepsilon)\sigma^2} - \frac{x}{y} \ln \left(1 + \frac{\alpha xy}{\varepsilon \sigma^2}\right) = 0.$$

Moreover, $\alpha_x$ incr. from $\varepsilon$ to $1$ as $x$ incr. from $0$ to $\infty$.

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

**Proposition**

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

$$\text{Pin}(x) \leq \text{PU}(x) = (\varepsilon + o(1))^{x/y} \text{Be}(x) \leq (\varepsilon + o(1))^{x/y} \text{BH}(x)$$

as $x \to \infty$.

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

---

Outline:
- Introduction
- Main results
- Sketch of proof
- Computation of bounds

Comparison of bounds
Here, $\sigma$ is normalized to be 1. In the next 4 frames, the graphs $G(P) := \{(x, \log_{10} \frac{P(x)}{\text{BH}(x)}): 0 < x \leq x_{\text{max}}\}$ for $P = \text{Ca}, \text{PU}, \text{Be}, \text{Pin}$, with the benchmark BH, will be shown, for $\varepsilon \in \{0.1, 0.9\}$, $y \in \{0.1, 1\}$, and $x_{\text{max}} = 3$ 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 $x_{\text{max}}$, the values of BH($x_{\text{max}}$) $\approx 0.016$ or 0.017, whether $y = 0.1$ or $y = 1$.

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

For $y = 1$, $\text{Ca}(x) < \text{BH}(x)$ 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: $x \in [0, 4]$, $\varepsilon = 0.1$, $y = 1$

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 $\text{Be}(x)$ is about 9.93 times worse (i.e., greater) than $\text{Pin}(x)$ at $x = 4$. Moreover, for these values of $\varepsilon$ and $y$, even the bound $\text{PU}(x)$ is better than $\text{Be}(x)$ already at about $x = 2.5$. 

$$
\text{BH} \sim 0 \\
\text{(BH}(4) \approx 0.017) \\
\text{Ca} \\
\text{Be} \\
\text{PU} \\
\text{Pin}
$$
Comparison: $x \in [0, 3], \varepsilon = 0.1, y = 0.1$

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 $\text{Be}(x)$ is still about 20% greater than $\text{Pin}(x)$ at $x = 3$. 

$\text{BH} \sim 0$  
$\text{BH}(4) \approx 0.016$  
$\text{Ca}$  
$\text{Be}$  
$\text{PU}$  
$\text{Pin}$
Comparison: $x \in [0, 4]$, $\varepsilon = 0.9$, $y = 1$

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 $\text{Be}(x)$ is about 8% better than $\text{Pin}(x)$ at $x = 4$. For $x \in [0, 4]$, $\text{Pin}(x)$ and $\text{Be}(x)$ are close to each other and both are significantly better than either $\text{BH}(x)$ or $\text{PU}(x)$ (which latter are also close to each other).
Comparison: $x \in [0, 4]$, $\varepsilon = 0.9$, $y = 0.1$

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 $\text{Be}(x)$ is about 12% better than $\text{Pin}(x)$ at $x = 3$. For $x \in [0, 3]$, $\text{Pin}(x)$ and $\text{Be}(x)$ are close to each other and both are significantly better than either $\text{BH}(x)$ or $\text{PU}(x)$ (which latter are very close to each other).
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: $y = 1$: distr. of the $X_i$’s may be much skewed to the right
Column 2: $y = 0.1$: distr. of the $X_i$’s may be only a little skewed to the right.
Thank you!