ESAIM: PS
September 2006, Vol. 10, p. 269–276
DOI: 10.1051/ps:2006011

ESAIM: Probability and Statistics
www.edpsciences.org/ps

STEIN ESTIMATION FOR INFINITELY DIVISIBLE LAWS *

R. Averkamp[1] and C. Houdré[2]

Abstract. Unbiased risk estimation, à la Stein, is studied for infinitely divisible laws with finite second moment.

Mathematics Subject Classification. 62G07, 62C20, 60G70, 41A25.

Received June 6, 2005. Revised November 18, 2005.

Let us start by briefly recalling the framework and results of Stein [6]: Let X_i, $i = 1, \dots, n$, be i.i.d. $N(0, \sigma^2)$ random variables and let $g = (g_1, \dots, g_n) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, be "weakly differentiable." Then for all $\theta \in \mathbb{R}^n$,

$$E\|X + \theta + g(X + \theta) - \theta\|_2^2 = n\sigma^2 + E\|g(X + \theta)\|_2^2 + 2\sigma^2 E \sum_{i=1}^{n} \frac{\partial}{\partial x_i} g_i(X + \theta), \tag{1}$$

where $\|\cdot\|_2$ is the Euclidean norm. Thus the risk of the estimator $x + g(x)$ can be estimated unbiasedly by $n\sigma^2 + g(x)^2 + 2\sigma^2 \sum_{i=1}^{n} \frac{\partial g_i}{\partial x_i}(x)$. This estimate is useful only if the variance of the risk estimate is small compared to the actual risk. This is especially the case if g_i only depends on X_i, since then the strong law of large numbers kicks in. For normal random variables the existence of the above estimates is based on the identity

$$\int_{\mathbb{R}} g'(x) \mathrm{e}^{-x^2/2} dx = \int_{\mathbb{R}} x g(x) \mathrm{e}^{-x^2/2} \mathrm{d}x.$$

Below, we obtain a corresponding identity for infinitely divisible random variables with finite variance, by replacing g' with $K(g)$ where K is an operator commuting with translations.

Let f be a density on $\mathbb{R}$, with mean 0 and variance σ^2 (for simplicity of notation we concentrate on the univariate case, but see Rem. 5). Let $d(x) = x + g(x)$ be an estimator in the location model induced by f. Let $F = \{f * \delta_\theta : \theta \in \mathbb{R}\}$, and let $L^2(F)$ and $L^1(F)$ have their canonical meaning. We want to estimate the risk of d unbiasedly:

$$\int_{\mathbb{R}} (d(x + \theta) - \theta)^2 f(x) dx = \int_{\mathbb{R}} g(x + \theta)^2 f(x) dx + \int_{\mathbb{R}} x^2 f(x) dx + 2 \int_{\mathbb{R}} x g(x + \theta) f(x) dx.$$

Keywords and phrases. Wavelets, thresholding, minimax.

* *Research supported in part by the NSF Grant DMS-9632032, while this author was visiting the Southeast Applied Analysis Center, School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA.*

[1] Institut für mathematische Stochastik Freiburg University Eckerstraße 1, 79104 Freiburg, Germany.

[2] School of Mathematics Georgia Institute of Technology Atlanta, GA 30332, USA; houdre@math.gatech.edu

In the above right hand side, the first summand can be estimated unbiasedly, the second is a constant, so we just need to find a function $h \in L^1(F)$ such that

$$\int_{\mathbb{R}} h(x+\theta)f(x)\mathrm{d}x = \int_{\mathbb{R}} xg(x+\theta)f(x)\mathrm{d}x. \tag{2}$$

If g is a polynomial the right-hand side of (2) is itself a polynomial in θ. It is then well-known that there exists an h satisfying (2). But if g is the soft-thresholding operator, *i.e.*, $g(x) = T_\lambda^S(x) = (|x| - \lambda)^+ \mathrm{sgn}(x)$, then g does not even have a power series expansion. Moreover, h does not have to be unique. Indeed, $h + q$ is also a solution for any function q such that $q * f = 0$ (if $\widehat{f}$, the Fourier transform of f, vanishes such q might exists).

Hence, let us assume that $\widehat{f}$ does not have any zero. By computing the generalized Fourier transform of both sides of

$$\int_{\mathbb{R}} g(-x+\theta)(-x)f(-x)\mathrm{d}x = \int_{\mathbb{R}} h(-x+\theta)f(-x)\mathrm{d}x, \tag{3}$$

we get:

$$\widehat{g}(w)(\widehat{f})'(-w) = i\widehat{h}(w)\widehat{f}(-w). \tag{4}$$

This identity shows that if $\widehat{g}$ converges to 0 fast enough, *e.g.*, if $\widehat{g}$ has compact support, then there exists an h such that (2) holds. Since $\widehat{f}$ does not vanish, h is uniquely determined. Hence the set

$$U_f := \left\{ g \in L^2(F) : \exists h \in L^1(F), \int_{\mathbb{R}} h(x+\theta)f(x)\mathrm{d}x = \int_{\mathbb{R}} xg(x+\theta)f(x)\mathrm{d}x, \forall \theta \in \mathbb{R} \right\},$$

is a vector space and clearly there is a unique linear map $K_f : U_f \longrightarrow L^1(f)$ with

$$\int_{\mathbb{R}} K_f(g)(x+\theta)f(x)\mathrm{d}x = \int_{\mathbb{R}} g(x+\theta)xf(x)\mathrm{d}x.$$

Let us present some properties of K_f:

Theorem 1. *Let f, f_1, f_2 be densities with finite second moment, and let K_f, K_{f_1} and K_{f_2} be well defined. Then for all $b \in \mathbb{R}$ and $g \in U_f$, respectively $g \in U_{f_1 * f_2}$:*

(1) $K_f(g(\cdot + b)) = K_f(g)(\cdot + b)$;
(2) $K_{f*\delta_b}(g) = K_f(g) + bg(\cdot)$;
(3) $K_{f_1 * f_2}(g) = K_{f_1}(g) + K_{f_2}(g)$;
(4) $K_{bf(\cdot b)}(g) = K_f(g(\cdot/b))(\cdot b)/b$, *for* $b > 0$.

Proof.
1. $\int_{\mathbb{R}} g(x+\theta+b)xf(x)\mathrm{d}x = \int_{\mathbb{R}} K_f(g)(x+\theta+b)f(x)\mathrm{d}x$ and thus $K_f(g(\cdot + b)) = K_f(g)(\cdot + b)$.
2.

$$\begin{aligned}\int_{\mathbb{R}} xg(x+\theta)f(x-b)\mathrm{d}x &= \int_{\mathbb{R}} (K_f(g)(x+\theta+b) + bg(x+\theta+b))f(x)\mathrm{d}x \\ &= \int_{\mathbb{R}} (K_f(g)(x+\theta) + bg(x+\theta))f(x-b)\mathrm{d}x.\end{aligned}$$

3. Let h_1, h_2 be such that $\int_{\mathbb{R}} g(x+\theta)xf_i(x)\mathrm{d}x = \int_{\mathbb{R}} h_i(x+\theta)f_i(x)\mathrm{d}x$, $i=1,2$, then

$$\begin{aligned}
&\int_{\mathbb{R}} (h_1+h_2)(z+\theta)(f_1 * f_2)(z)\mathrm{d}z \\
&\quad = \int_{\mathbb{R}}\int_{\mathbb{R}} (h_1(x+y+\theta)+h_2(x+y+\theta))f_1(x)f_2(y)\mathrm{d}x\mathrm{d}y \\
&\quad = \int_{\mathbb{R}}\int_{\mathbb{R}} h_1(x+y+\theta)f_1(x)\mathrm{d}xf_2(y)\mathrm{d}y + \int_{\mathbb{R}}\int_{\mathbb{R}} h_2(y+x+\theta)f_2(y)\mathrm{d}yf_1(x)\mathrm{d}x \\
&\quad = \int_{\mathbb{R}} g(z+\theta)z(f_1 * f_2)(z)\mathrm{d}z.
\end{aligned}$$

4.

$$\begin{aligned}
\int_{\mathbb{R}} g(x+\theta)x(bf(bx))\mathrm{d}x &= \int_{\mathbb{R}} g(x/b+\theta)xf(x)/b\mathrm{d}x \\
&= \int_{\mathbb{R}} g((x+b\theta)/b)x/bf(x)\mathrm{d}x \\
&= \int_{\mathbb{R}} K_f(g(\cdot/b))(x+b\theta)/bf(x)\mathrm{d}x \\
&= \int_{\mathbb{R}} K_f(g(\cdot/b))((x+\theta)b)/b(bf(xb))\mathrm{d}x, \qquad \square
\end{aligned}$$

and thus $K_{bf(\cdot b)}(g) = K_f(g(\cdot/b))(\cdot b)/b$.

Note that the third property presented above is very useful for wavelet analysis, since the law of the noise in a wavelet coefficient is a weighted convolution of the noise in the original data.

For the normal distribution with unit variance the operator K is given by $K(g)=g'$, *i.e.* K is the differentiation operator. In general K is quite complicated to compute, however from (4) we see that formally

$$\widehat{K_f(g)}(w) = \widehat{g}(w)(\widehat{f})'(-w)/(\widehat{f}(-w)i).$$

This suggest that h can be computed by a convolution of the estimator and of a function or measure, which is the inverse Fourier transform of $(\widehat{f})'(-w)/(\widehat{f}(-w)i)$. Let us try to further formalize this claim. Assume that $K_f(g) := K_f * g$, where $K_f \in L^1(\mathbb{R})$ and $\widehat{K}_f = (\widehat{f})'(-\cdot)/(\widehat{f}(-\cdot)i)$. If $g \in L^\infty(\mathbb{R})$, then $K_f * g$ does what it is supposed to do:

$$\begin{aligned}
\int_{\mathbb{R}} (K_f * g)(x+\theta)f(x)\mathrm{d}x &= \int_{\mathbb{R}}\int_{\mathbb{R}} K_f(x-t)g(t+\theta)\mathrm{d}tf(x)\mathrm{d}x \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} K_f(x-t)f(x)\mathrm{d}xg(t+\theta)\mathrm{d}t \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} K_f(-(t-x))f(x)\mathrm{d}xg(t+\theta)\mathrm{d}t \\
&= \int_{\mathbb{R}} (K_f(-\cdot) * f(\cdot))(t)g(t+\theta)\mathrm{d}t \\
&= \int_{\mathbb{R}} tf(t)g(t+\theta)\mathrm{d}t,
\end{aligned}$$

where the last equality follows from the construction of K:

$$\widehat{K}(-\cdot)\widehat{f} = \frac{(\widehat{f})'}{i} = (\widehat{f(\cdot)\mathrm{id}}).$$

If K_f is known, then it can still be a problem to compute $K_f(g)$, since $K_f(g)$ is not necessarily as simple as g'. If $g = \sum_i g_i$ and the $K_f(g_i)$ are easy to compute, then we can compute $K_f(g)$ since K_f is linear. For example we can take $g_\lambda^+(x) = (x-\lambda)^+$ and $g_\lambda^-(x) := (x-\lambda)^-$ as simple building blocks for functions. Note that $K_f(g_\lambda^+)(x) = K_f(g_0^+)(x-\lambda)$ and $K_f(g_0^-)(x) = \sigma^2 - K_f(g_0^+)(x)$, if $\int_{\mathbb{R}} xf(x)\mathrm{d}x = 0$ and $\int_{\mathbb{R}} x^2 f(x)\mathrm{d}x = \sigma^2$. For example the soft thresholding estimator T_λ^S as well as T_λ^M given by $T_\lambda^M(x) := x\mathbf{1}_{\{|x|\geq\lambda\}} + 2(|x|-\lambda/2)_+\mathrm{sgn(x)}\mathbf{1}_{\{|x|<\lambda\}}$, have the following decompositions:

$$T_\lambda^S(x) = x - g_0^+(x) + g_\lambda^+(x) - g_0^-(x) + g_{-\lambda}^-(x), \tag{5}$$

$$T_\lambda^M(x) = x - g_0^+(x) + 2g_{\lambda/2}^+(x) - g_\lambda^+(x) - g_0^-(x) + 2g_{-\lambda/2}^-(x) - g_{-\lambda}^-(x).$$

For a further example, assume that $g : \mathbb{R}^+ \longrightarrow \mathbb{R}$, is twice continuously differentiable with $g(0) = 0$, then $g(x) = g'(0^+)x^+ + \int_0^\infty (x-y)^+ g''(y)\mathrm{d}y$.

Another simple example is provided by compound Poisson distributions. Indeed, let F be a compound Poisson distribution with Fourier transform $\exp(\lambda(\Psi(w)-1))$, where Ψ is the characteristic function of the density f. Then

$$\widehat{K}_F = \frac{\lambda\Psi'(-w)}{i},$$

and thus $K_F(x) = -\lambda f(-x)x$, *i.e.* $K_F(g) = K_F * g$. Since compound distributions are building blocks for infinitely divisible ones, and since infinitely divisible characteristic functions have no real zeros, we are led to:

Theorem 2. *Let f be an infinitely divisible density with finite second moment, i.e., let*

$$\widehat{f}(t) = \exp\left(ibt + \int_{\mathbb{R}} \left(\frac{\exp(ixt) - 1 - ixt}{x^2}\right) M(\mathrm{d}x)\right), \tag{6}$$

where M, the Lévy measure, is a finite positive measure. Let $M(\{0\}) = 0$, let $b = 0$ and let g be Lipschitz. Then,

$$K(g)(t) := \int_{\mathbb{R}} \frac{g(t+x) - g(t)}{x} M(\mathrm{d}x), \tag{7}$$

is a well defined real valued function, which is moreover bounded and continuous. Furthermore,

$$\int_{\mathbb{R}} K(g)(x+\theta)f(x)\mathrm{d}x = \int_{\mathbb{R}} xg(x+\theta)f(x)\mathrm{d}x.$$

Proof. It is clear that $K(g)$ is well defined, bounded and continuous. If g has compact support then $\int_{\mathbb{R}} |g(x+y) - g(x)|/|y|M(\mathrm{d}y)$ and $K(g)$ are in $L^1(\mathbb{R})$ and

$$\begin{aligned}
\widehat{K(g)}(t) &= \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{g(y+x) - g(y)}{x} M(\mathrm{d}x)\exp(ity)\mathrm{d}y \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{g(y+x) - g(y)}{x} \exp(ity)\mathrm{d}yM(\mathrm{d}x) \\
&= \widehat{g}(t)\int_{\mathbb{R}} \frac{\exp(-ixt) - 1}{x} M(\mathrm{d}x).
\end{aligned}$$

Since $\int_{\mathbb{R}}(\exp(-ixt)-1)/xM(\mathrm{d}x) = (\widehat{f})'(-t)/(\widehat{f}(-t)i)$, the Fourier transforms of $\int_{\mathbb{R}} K(g)(x+\theta)f(x)\mathrm{d}x$ and of $\int_{\mathbb{R}} g(x+\theta)xf(x)\mathrm{d}x$ are equal and thus these two terms are themselves equal, for all $\theta \in \mathbb{R}$.

If g is Lipschitz but does not have compact support, then let $g_n(x) := (1 - |x|/n)_+ g(x)$. Then the Lipschitz constants of the g_n form a bounded set. Moreover, g_n and $K(g_n)$ converge pointwise, respectively, to g and to $K(g)$. Clearly, since $\|K(g_n)\|_\infty$ is bounded, $\lim_{n\longrightarrow\infty} \int_{\mathbb{R}} g_n(x+\theta)xf(x)\mathrm{d}x = \int_{\mathbb{R}} g(x+\theta)xf(x)\mathrm{d}x$, while $\lim_{n\longrightarrow\infty} \int_{\mathbb{R}} K(g_n)(x+\theta)f(x)\mathrm{d}x = \int_{\mathbb{R}} K(g)(x+\theta)f(x)\mathrm{d}x$, for all θ. Hence,

$$\int_{\mathbb{R}} K(g)(x+\theta)f(x)\mathrm{d}x = \int_{\mathbb{R}} g(x+\theta)xf(x)\mathrm{d}x.$$

□

Remark 3. The assumption $b = 0$, is not serious, b is a location parameter of the density and thus we can appeal to Theorem 1. The condition $M(\{0\}) = 0$ is not restrictive either. If $M(\{0\}) = \sigma^2$, then the distribution is the convolution of a centered normal distribution with variance σ^2 and of an infinitely divisible distribution with Lévy measure without atom at the origin. Again we can use Theorem 1 in this situation. Hence in the general case we obtain:

$$K(g)(t) := bg(t) + M(\{0\})g'(t) + \int_{\mathbb{R}\setminus\{0\}} \frac{g(t+x) - g(t)}{x} M(\mathrm{d}x).$$

We also note here that although of little interest to us since we are dealing with mean square errors, the operator K in (7) could as well be defined (slightly modified) just under a finite first moment assumption on X (in this case, the representation (6) should be slightly changed, and so should be the requirements on M).

Remark 4. Let f be a density with mean zero and variance σ^2. Without loss of generality let also $K_f(1) = 0$. Let $f_n = *_{i=1}^n \sqrt{n} f(\cdot\sqrt{n})$. By the central limit theorem f_n converges in distribution to a normal density. So one would expect K_{f_n} to converge in some sense to $\sigma^2\mathrm{d}/\mathrm{d}x$. Assume that $K_f(g)(x) = \int_{\mathbb{R}}(g(x+y) - g(x))/yM(\mathrm{d}y)$. Note that if $K_f(g) = Q * g$, where Q is a finite real measure, then with the notation $Q^-(A) := Q(-A)$,

$$\begin{aligned}(Q * g)(x) &= \int_{\mathbb{R}} (g(x-y) - g(x))\, Q(\mathrm{d}y) \\ &= \int_{\mathbb{R}} \frac{g(x+y) - g(x)}{y} yQ^-(\mathrm{d}y),\end{aligned}$$

where the first equality holds since $K_f(1) = 0$, *i.e.*, $\int_{\mathbb{R}} 1Q(\mathrm{d}x) = 0$. Next, since $\int_{\mathbb{R}} x(x+\theta)f(x)\mathrm{d}x = \int_{\mathbb{R}} x^2 f(x)\mathrm{d}x$, taking $g(x) = x := i\mathrm{d}(x)$ gives $K_f(x) := K_f(i\mathrm{d})(x) = M(\mathbb{R})$ and thus $\int_{\mathbb{R}} K_f(x)f(x)\mathrm{d}x = \int_{\mathbb{R}} x^2 f(x)\mathrm{d}x = \sigma^2$. But, from Theorem 1, we know that $K_{f_n}(g)(x) = nK_f(g(\cdot/\sqrt{n}))(x\sqrt{n})/\sqrt{n}$. Using the form of K_f, this gives:

$$K_{f_n}(g)(x) = \int_{\mathbb{R}} \frac{g(x + y/\sqrt{n}) - g(x)}{y/\sqrt{n}} M(\mathrm{d}y).$$

Thus, if g is Lipschitz and differentiable, $\lim_{n\longrightarrow\infty} K_{f_n}(g)(x) = \sigma^2 g'(x)$.

Examples 1. Let $f(x) = \exp(-\sqrt{2}|x|)/\sqrt{2}$ be the variance normalized Laplace density. It is easy to see that, $\widehat{f}(w) = 2/(2 + w^2)$. Thus

$$\frac{(\widehat{f})'(w)}{i\widehat{f}(w)} = \frac{2iw}{2 + w^2},$$

and

$$\widehat{K_f}(w) = \frac{-2iw}{2 + w^2} = -iw\widehat{f}(w) = \widehat{(f')}(w).$$

Thus $K_f(x) = -\exp(-\sqrt{2}|x|)\mathrm{sgn}(x) \in L^1(\mathbb{R})$. Tedious but simple computations yield

$$K_f * x_+ = \left\{ \begin{array}{ccc} \frac{\exp(\sqrt{2}x)}{2} & : & x \leq 0 \\ 1 - \frac{\exp(-\sqrt{2}x)}{2} & : & x > 0 \end{array} \right\} := h(x).$$

Using (5) we obtain

$$\begin{aligned} K_f(T_\lambda^S(x) - x) &= -h(x) - (1 - h(x)) + h(x - \lambda) + (1 - h(x + \lambda)) \\ &= h(x - \lambda) - h(x + \lambda). \end{aligned}$$

Combining these results, we see that if X has a Laplace distribution,

$$E(T_\lambda^S(X + \theta) - \theta)^2 = 1 + E\min((X + \theta)^2, \lambda^2) + 2(h(X + \theta - \lambda) - h(X + \theta + \lambda)).$$

2. Let $f_t(x) = \exp(-x)x^{t-1}/\Gamma(t)\mathbf{1}_{\mathbb{R}^+}(x)$ be the density of the Gamma distribution. Since the mean of this distribution is t we want to compute $K_{f_t * \delta_{-t}}$. Then by Feller [5], p. 567 $\log \widehat{f}_t(x) = t\int_0^\infty (\exp(iyx) - 1)/y \exp(-y)\mathrm{d}y$ and thus $(\log \widehat{f}_t)'(x) = it\int_0^\infty \exp(iyx)\exp(-y)\mathrm{d}y$. Hence $K_{f_t}(g) = Q * g$ where $Q \in L^1(\mathbb{R})$ is given via

$$\widehat{Q}(x) = t\int_0^\infty \exp(-iyx)\exp(-y)\mathrm{d}y = t\int_{-\infty}^0 \exp(ixy)\exp(y)\mathrm{d}y.$$

Hence $K_{f_t * \delta_{-t}}(g)(x) = t\int_{-\infty}^0 \exp(y)g(x - y)\mathrm{d}y - tg(x)$.

3. Another example is the cosine hyperbolic density, $f(x) = 1/\cosh(\pi x/2)$, again [5] p. 567

$$\log(\widehat{f}(x)) = \int_{\mathbb{R}} \frac{\exp(ixy) - 1 - iyx}{y^2} \frac{y}{\exp(y) - \exp(-y)}\mathrm{d}y$$

and thus

$$K_f(g)(x) = \int_{\mathbb{R}} \frac{g(x + y) - g(x)}{y} \frac{y}{\exp(y) - \exp(-y)}\mathrm{d}y.$$

All the examples presented above are infinitely divisible and so K has a relatively nice form. Let us consider a case which is not: The uniform distribution with density $2^{-1}\mathbf{1}_{(-1,1)}$. Assume $g : [-1, 1] \longrightarrow \mathbb{R}$ and $2^{-1}\int_{-1}^1 g(x)\mathrm{d}x = 0$. If $\bar{g}$ is the 2–periodic extension (to $\mathbb{R}$) of g then $2^{-1}\bar{g} * \mathbf{1}_{(-1,1)} = 0$. Thus unbiased risk estimators are not uniquely determined. Let

$$r(\theta) = 2^{-1}\int_{-1}^1 x(x + \theta)_+\mathrm{d}x = \begin{cases} 0 & : \theta \leq -1 \\ \frac{1}{6} + \frac{\theta}{4} - \frac{\theta^3}{12} & : \theta \in (-1, 1), \\ \frac{1}{3} & : \theta \geq 1. \end{cases}$$

After a few failed attempts one finds that with

$$h(x) = \begin{cases} 0 & : x \leq 0 \\ -\frac{(x - [x/2]2)(x - [x/2]2 - 2)}{2} & : x \geq 0. \end{cases}$$

(h is the 2-periodic extension, to $\mathbb{R}^+$, of $-x(x-2)/2$ which is defined on $[0, 2]$.) $2^{-1}\int_{-1}^1 h(x+\theta)\mathrm{d}x = r(\theta)$. So with the help of (5) we can now compute an unbiased risk estimator for soft thresholding. Figure 1 shows the unbiased risk estimators for soft thresholding with threshold 2 for the normal distribution, the Laplace distribution, the

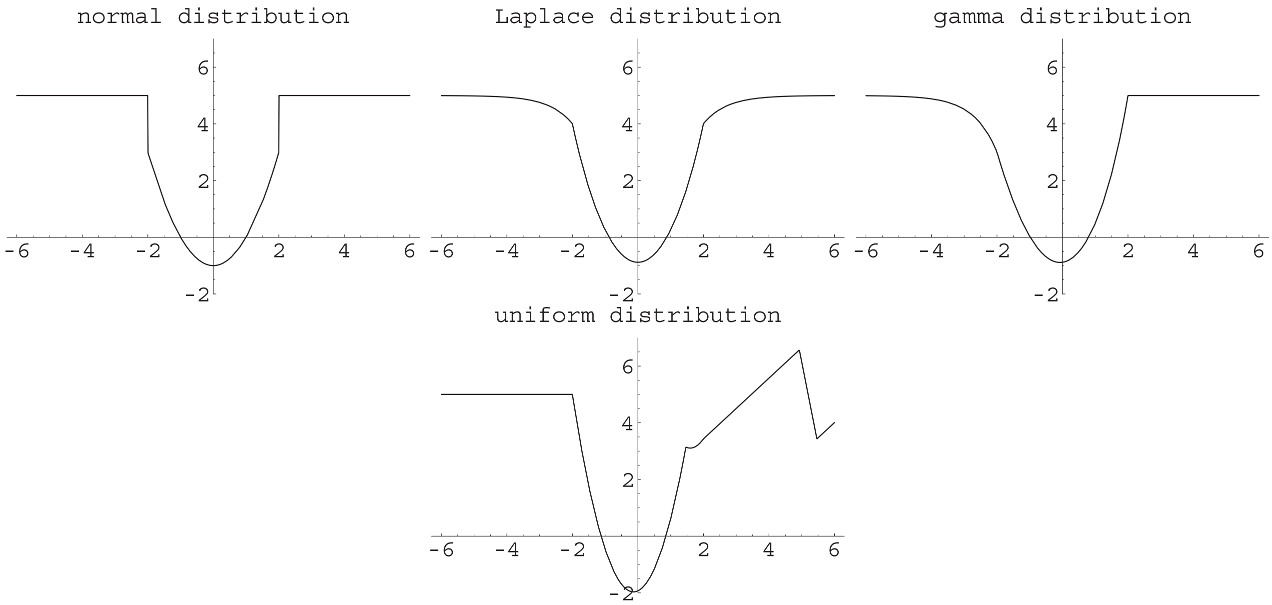

FIGURE 1. The unbiased risk estimators for soft thresholding.

gamma distribution with $t = 2$ and the uniform distribution. The distributions were transformed to have mean zero and variance one. The horizontal axis corresponds to the observed values while the vertical one gives the estimated risks.

Remark 5. As we have seen with (1), for normal random variables, unbiased risk estimation is possible for multivariate means, even if the estimators for the coordinates are not independent. This is also possible for other types of distributions, one has only to apply the operator K coordinatewise. Indeed, let X_i, $i = 1, \ldots, n$ be random variables such that X_i has distribution F_i and $EX_1 = 0$, $EX_1^2 = \sigma_1^2$. Assume that an operator K_1 exists such that $EX_1 g(X_1 + \theta_1) = EK_1(g)(X_1 + \theta_1)$ for some g. If $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $\theta \in \mathbb{R}^n$ then $E(X_1 + g(X + \theta) - \theta_1)^2 = \sigma_1^2 + Eg(X + \theta)^2 + 2EX_1 g(X + \theta)$. Then, under the proper conditions on g,

$$\begin{aligned} EX_1 g(X+\theta) &= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} x_1 g(x_1+\theta_1, \ldots, x_n+\theta_n) F_1(\mathrm{d}x) \otimes_{i=2}^n F_i(\mathrm{d}(x_2, \ldots, x_n)) \\ &= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} K_1(g(\cdot, x_2+\theta_2, \ldots, x_n+\theta_n))(x_1+\theta_1) F_1(\mathrm{d}x) \otimes_{i=2}^n F_i(\mathrm{d}(x_2, \ldots, x_n)). \end{aligned}$$

Thus $E(X_1 + g(X + \theta) - \theta_1)^2 = \sigma_1^2 + Eg(X + \theta)^2 + 2EK_1(g(\cdot, X_2 + \theta_2, \ldots, X_n + \theta_n)(X_1 + \theta_1)$.

Actually, the multivariate infinitely divisible case can be done at once. Multivariate unbiased risk estimation for $\theta := (\theta_1, \ldots, \theta_n)$ can be reduced to unbiased risk estimation for the coordinates $\theta_i \in \mathbb{R}$. The estimator for $\theta \in \mathbb{R}^n$ is given by $d(x) := x + g(x), x \in \mathbb{R}^n$, where $g := (g_1, \ldots, g_n) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$. Next, for the i-th component, set

$$K_i(g_i)(t) := \int_{\mathbb{R}^n} \frac{x_i(g_i(t+x) - g_i(t))}{\|x\|^2} M(\mathrm{d}x),$$

where now the Lévy measure M is a positive measure on $\mathbb{R}^n$, without atom at the origin. It then follows that

$$\int_{\mathbb{R}^n} K_i(g_i)(x+\theta) f(x) \mathrm{d}x = \int_{\mathbb{R}^n} x_i g_i(x+\theta) f(x) \mathrm{d}x,$$

and from there we define the operator $K := (K_1, \ldots, K_n)$.

Remark 6. As the reader might have figured out by now, the motivation for the present paper comes from thresholding methods in wavelet denoising (see [4]). In a function space approach to denoising, the thresholds depend on the sample size n, on the Besov space to which the target functions belong to, and also on the Besov norm of these targets. In practice it is often not known which threshold is appropriate since the function space to which the signal belongs as well as the value of its norm are unknown. To bypass this problem, Donoho and Johnstone developed a procedure called SureShrink where thresholds are chosen automatically (see [3]). Their method, based on Stein's unbiased risk estimate is as follows: for each level (except the highest levels) in the noisy wavelet transform, the largest threshold (smaller than $\sqrt{2\log n}$) which minimizes the unbiased risk estimate is chosen. For soft thresholding finding this minimum is simple and takes $O(n\log n)$ time.

As noticed in [1], the central limit theorem works fast for wavelet coefficients, so it is reasonable to apply the normal adaptive results to the general non Gaussian framework. However, it is also of interest to understand the scope of SureShrink beyond the normal framework. To do so, we needed to find unbiased risk estimates for other types of distributions. This is what we did here for infinitely divisible and related noise. Indeed, the (noisy) wavelet coefficients are linear combinations of the (noisy) input signal. Hence, if we are able to compute the unbiased risk estimator, *i.e.*, the operator K for the original (input) noise, then with the help of Theorem 1 this can also be done for the noise in the wavelet coefficients. However, unlike in the Gaussian case, the coordinatewise unbiased risk estimators are no longer independent. Also, our methods enable us to compute the bias when applying the unbiased risk estimator for Gaussian noise to other kind of noise. Remark 4 gives some hints that this bias becomes smaller for higher level, *i.e.*, coarser, wavelet coefficients. With the approach just briefly described, the scope of SureShrink could then be potentially extended using also the corresponding threshold levels found in [1] and [2].

References

[1] R. Averkamp and C. Houdré, Wavelet Thresholding for non necessarily Gaussian Noise: Idealism. *Ann. Statist.* **31** (2003) 110–151.

[2] R. Averkamp and C. Houdré, Wavelet Thresholding for non necessarily Gaussian Noise: Functionality. *Ann. Statist.* **33** (2005) 2164–2193.

[3] D.L Donoho and I.M. Johnstone, Adapting to Unknown Smoothness *via* Wavelet Shrinkage. *J. Amer. Statist. Assoc.* **90** (1995) 1200–1224.

[4] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian and D. Picard, Wavelet Shrinkage: Asymptotia? *J. Roy. Statist. Soc. Ser. B* **57** (1995) 301–369.

[5] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II. John Wiley & Sons (1966).

[6] C. Stein, Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** (1981) 1135–1151.