Information Cost Functions*

Mladen Victor Wickerhauser†

Department of Mathematics
Washington University
St. Louis, Missouri, 63130 USA

`victor@math.wustl.edu`
`www.math.wustl.edu/~victor`

Zhejiang University
27 May 2004

*Research supported by NSF, AFOSR, and SWB, Inc.
†Joint work with Hrvoje Šikić

Approximate a Signal

A good approximation basis concentrates the signal's energy:

- only a relatively tiny number of components are non-negligible;

- individually negligible coefficients add up to a negligible sum.

Idea: Component amplitudes, when rearranged into decreasing order, decrease rapidly.

Problem: How to compare rates of decrease and find a good basis effficiently.

Solution: Use a divide-and-conquer strategy to search a large number of bases. But it is necessary to avoid rearrangement in tests of basis quality.

From Basis to PDF

Let B be an orthonormal basis for a separable Hilbert space X with norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$, and let $x \in X$ be a fixed vector.

Expand $x' = x/\|x\|$ in $B = \{b_k\} \in \mathbf{B}$ to get a coefficient sequence $c_k^B = \langle x', b_k\rangle$, and define

$$p_k^B = |c_k^B|^2.$$

Then $p^B = p^B(x)$ is a discrete probability density function, or pdf.

The rate at which best N-component approximations converge to x, as $N \to \infty$, is related to the concentration of the pdf $p^B(x)$.

Rearranged Partial Sums

Definition 1 *The (rearranged) partial sum of a pdf p is the sequence*

$$Sp_n = \sum_{k \le n} p_k^*,$$

where p^* is the decreasing rearrangement of p.

The sequence Sp increases monotonically to 1.

Sp is concave in the sense that

$$Sp_{k-1} + Sp_{k+1} \le 2Sp_k,$$

whenever all indices are defined.

Comparing Rates of Approximation

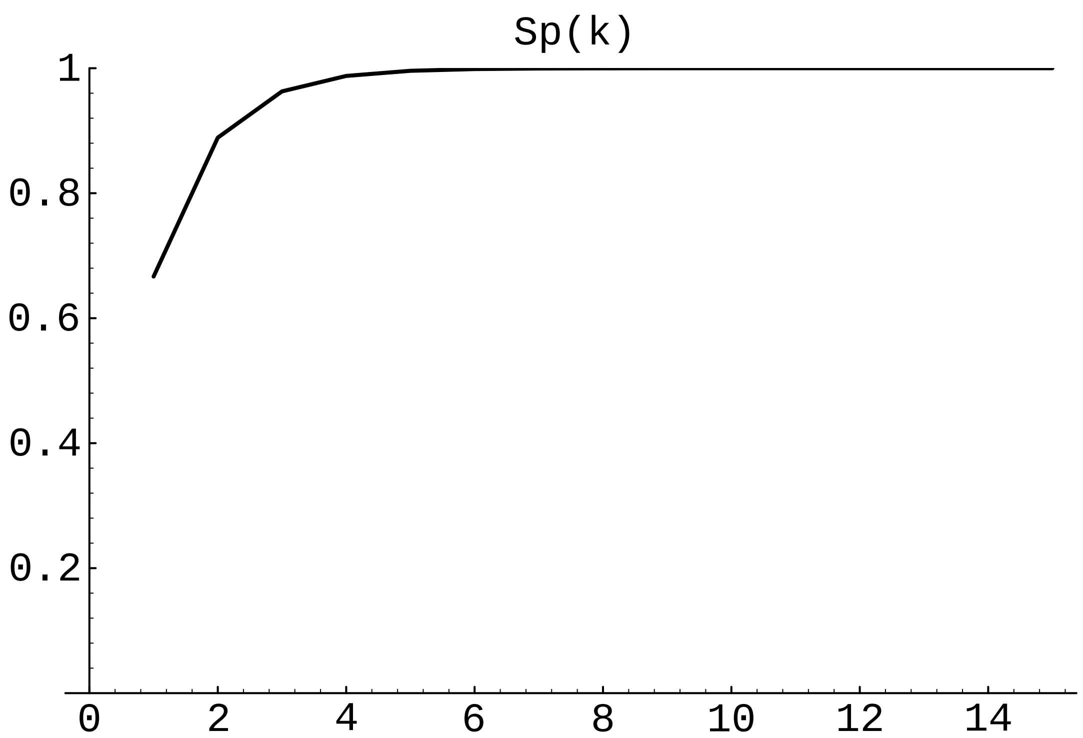
Definition 2 *For pdfs p, q, we say that p majorizes q, and write $p \succ q$, if $(\forall n) Sp_n \geq Sq_n$.*

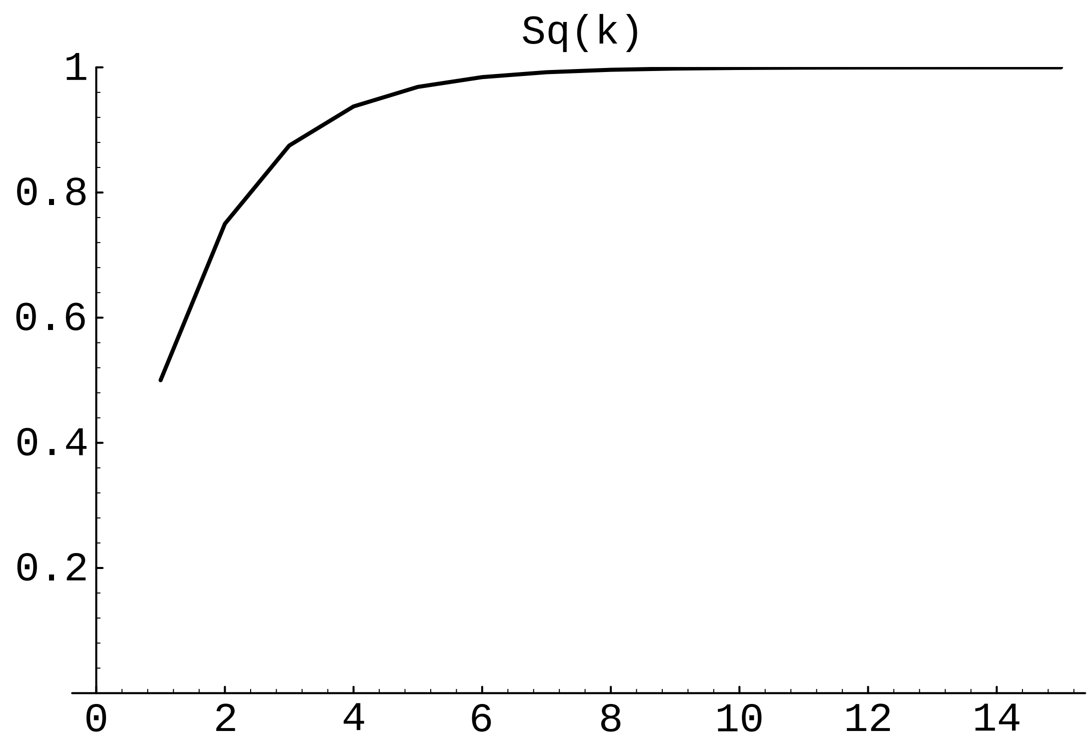
Suppose X is a separable Hilbert space with norm $\|\cdot\|$, $x \in X$, and B, B' are orthonormal bases for X.

If $p^B(x) \succ p^{B'}(x)$, then for all N, the best N-component approximation of x in basis B is at least as close as the best N-component approximation of x in B'.

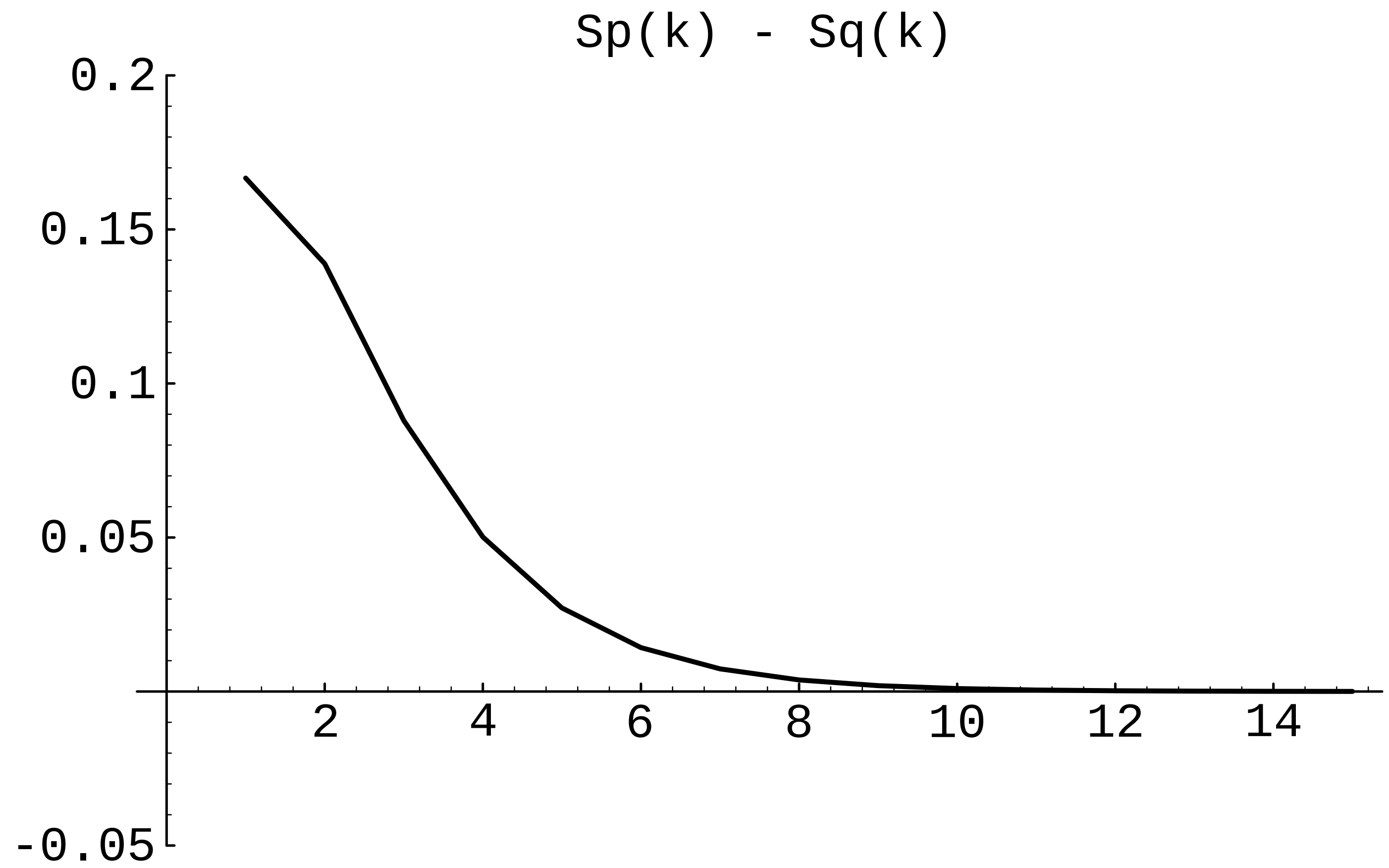
Majorization Examples

Example: $p_k = 2 \times 3^{-k}$ majorizes $q_k = 2^{-k}$.





Majorization Examples (continued)



Proof: $Sp_k - Sq_k = (1 - 3^{-k}) - (1 - 2^{-k}) > 0$.

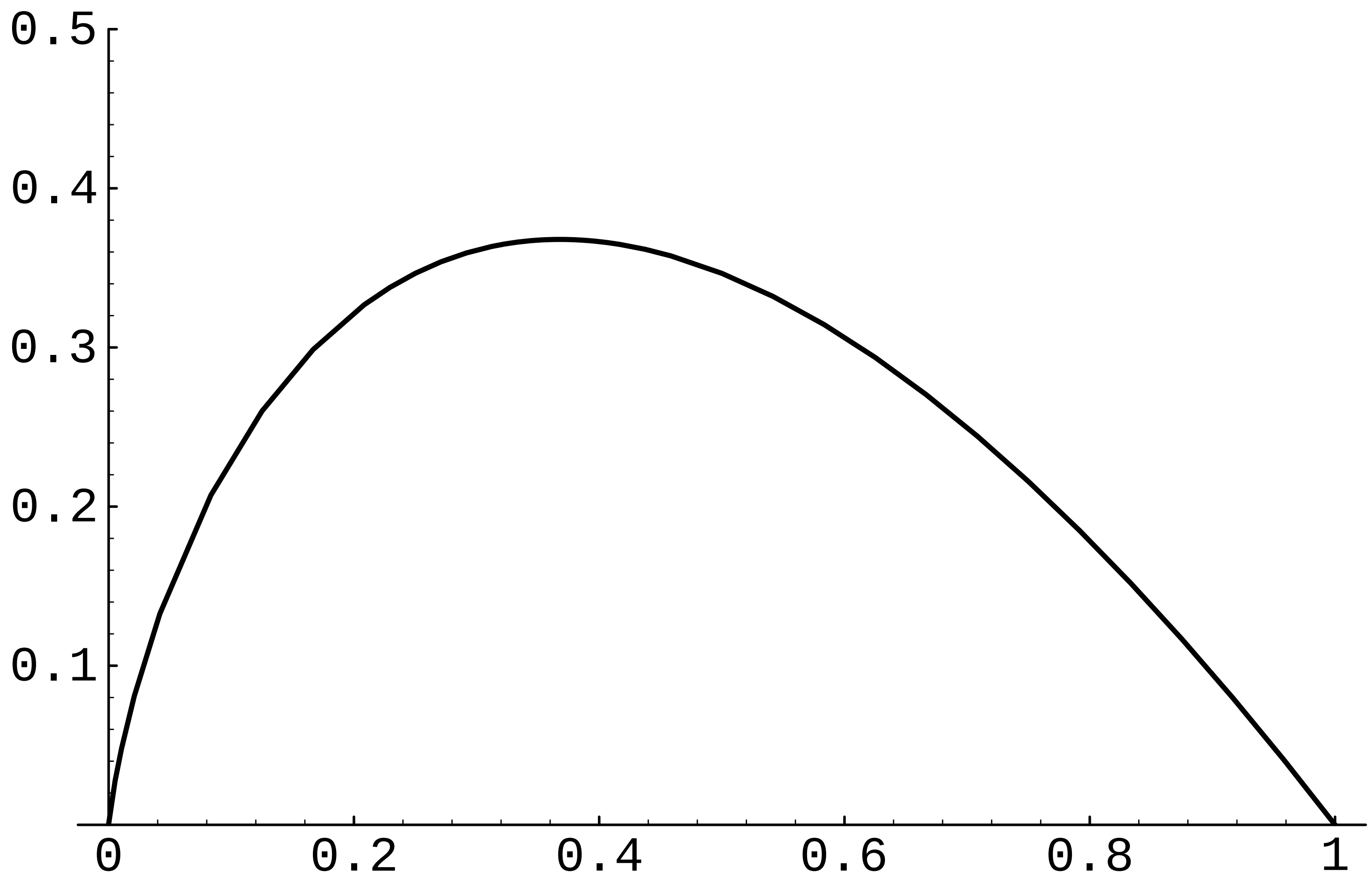
Additive Information Cost

Definition 3 *An additive information cost function H is a nonnegative concave Schur functional on pdfs, namely, a functional of the form*

$$H(p) = \sum_k h(p_k),$$

where $h \geq 0$ is a concave function on $[0, 1]$.

Example: entropy functional $h(t) = -t \log t$.



The Best Basis in a Library

A library $\mathbf{B}$ is a collection of orthonormal bases for a separable Hilbert space X.

For a fixed vector $x \in X$, let $p^B = p^B(x)$ be the pdf associated to the expansion of x in the basis $B \in \mathbf{B}$.

Let H be an additive information cost function.

Definition 4 *The best basis $B \in \mathbf{B}$ for a given x and H is the minimizer of $H(p^B)$.*

The Fastest Approximation Basis

Definition 5 *The fastest approximation basis $B \in \mathbf{B}$ for a given x is one that satisfies*

$$p^B(x) \succ p^{B'}(x),$$

for all $B' \in \mathbf{B}$.

Only in very special cases of $\mathbf{B}$ and x is it known whether a fastest approximation basis exists.

But the best basis for any single information cost function is the sole candidate for fastest approximation basis:

Theorem 1 (Hardy-Littlewood-Pólya, 1929) *Suppose p, q are finitely-supported pdfs. Then $p \succ q$ if and only if $H(p) \leq H(q)$ for every additive information cost function H.*
$\square$

Three Results

We can estimate rates of decrease, avoiding rearrangement and summation, using only information cost functions:

- Hardy–Littlewood–Pólya(1929) is true for infinite sequences.

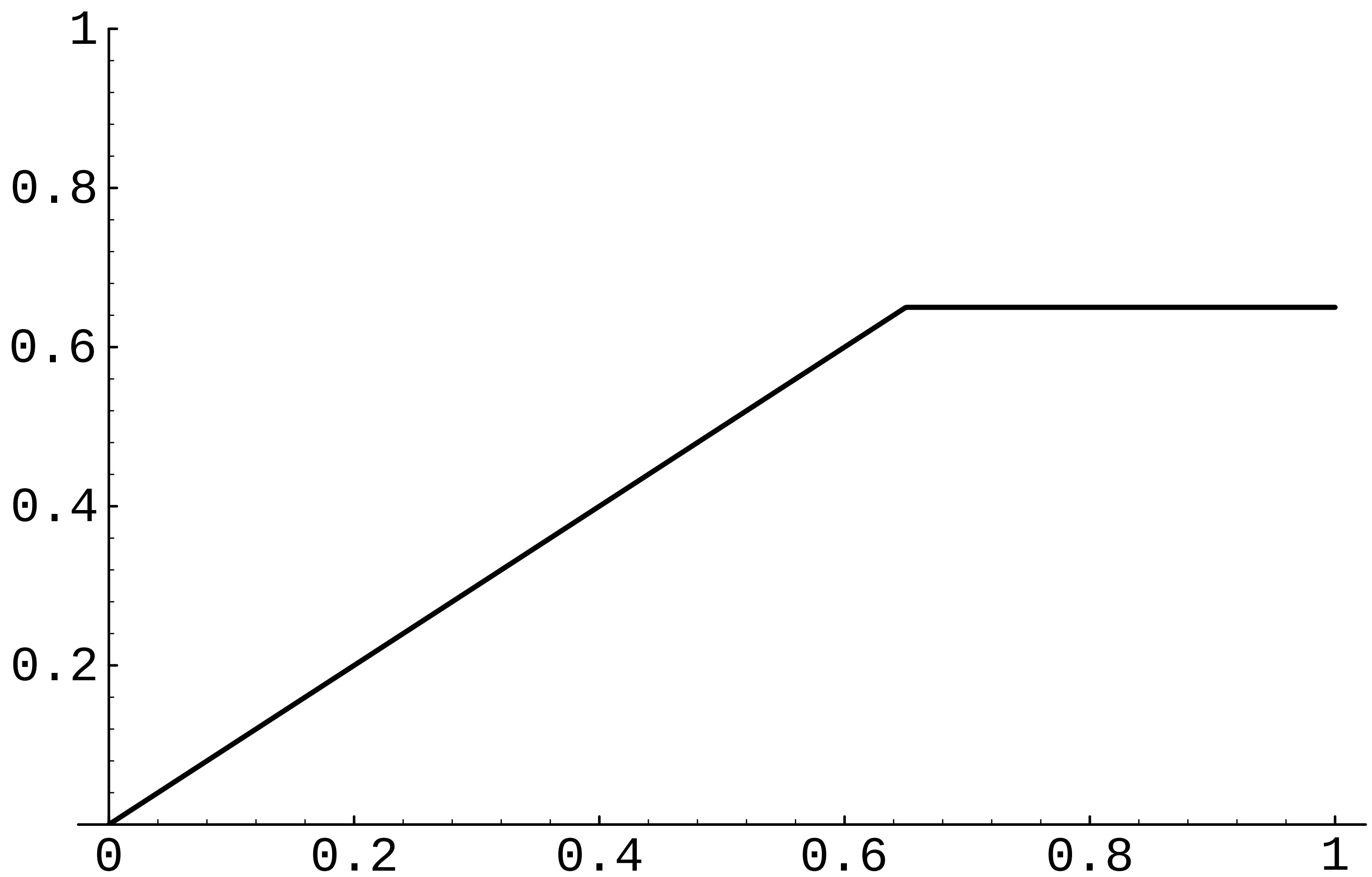
- Sub-exponential $h(t)/t$ satisfy inequalities relating different information costs.

- Information cost functions are comparable: the cost of a pdf is bracketed between two simpler functions of the pdf

Infinite Version of H–L–P(1929)

Theorem 2 *Suppose p and q are discrete pdfs. Then $p \succ q$ if and only if $H(p) \leq H(q)$ for every additive information cost function H.* □

In fact, the "if" part holds for the class of information costs generated by the set of threshold functions $\{h_T : T > 0\}$, defined by $h_T(t) = \min\{t, T\}$.



The example above is $h_{0.65}(t)$.

Subexponentiality

Fix $d > 0$.

Definition 6 *A function $\ell : (0,1) \to \mathbf{R}$ is called d-subexponential if*

$$\ell(t^d) \leq d\ell(t),$$

for all $0 < t < 1$.

Definition 7 *The $(-d)$-subexponential extension of a function $\ell : (0,1) \to \mathbf{R}$ is the function $\ell : (1,\infty) \to \mathbf{R}$ defined by*

$$\ell(t) = -d\ell(t^{-\frac{1}{d}}),$$

for all $t > 1$.

Inequalities for Information Costs

For additive information cost function H with concave nonnegative h, define $\ell(t) = h(t)/t$. Then ℓ is nonnegative. Let p be a pdf.

Proposition 3 *If ℓ is nonnegative and convex on $(0,1)$, then $0 \le \ell\left(\sum_n p_n^2\right) \le H(p)$. Equality holds on the right if $p_n = 1$ for a single n.* □

Proposition 4 *Fix $d > 0$. If ℓ is nonnegative, convex on $(0,1)$, and d-subexponential, then*

$$0 \le \frac{1}{d}\ell\left(\sum_n p_n^{1+d}\right) \le H(p),$$

for any pdf p. □

Proposition 5 *Fix $d > 0$. If ℓ is convex on $(1,\infty)$ and $(-d)$-subexponential, then*

$$H(p) \le -\frac{1}{d}\ell\left(\sum_n p_n^{1-d}\right),$$

for any pdf p such that both $\{p_n^{1-d}\}$ and $\{p_n \ell(p_n^{-d})\}$ are summable. □

Application: An Inequality

Fix $0 < \alpha < 1$; put $\ell(t) = t^{-\alpha} - 1$ for $t \in (0, 1)$.

This ℓ is nonnegative, decreasing, convex on $(0, 1)$, and d-subexponential for all $d \in (0, 1)$. Applying the propositions, we can prove:

Lemma 6 *For* $0 < \alpha \leq d < 1$ *and every pdf* p *for which* $\{p_n^{1-d}\}$ *is summable, the following inequality holds:*

$$\left(\sum_n p_n^{1-\alpha}\right)^d \leq \left(\sum_n p_n^{1-d}\right)^\alpha.$$

Legendre Transforms

Definition 8 *The Legendre transform of a function* $\ell : I \to \mathbf{R}$ *is the function*

$$\tilde{\ell}(t) = \text{sup}\{at + b : \forall s \in I, as + b \leq \ell(s)\}.$$

Lemma 7 *Either* $\tilde{\ell}(t) = -\infty$ *for all* $t \in I$*, or else* $\tilde{\ell}$ *is finite and convex on* I*, and satisfies* $-\infty \leq \tilde{\ell}(t) \leq \ell(t)$ *for all* $t \in I$*. Furthermore,* $\tilde{\ell}$ *is the greatest convex function below* ℓ*, in the sense that if* $c = c(t)$ *is convex and* $c(t) \leq \ell(t)$ *for all* $t \in I$*, then* $c(t) \leq \tilde{\ell}(t)$ *for all* $t \in I$*.* □

Note that if ℓ is convex, then $\tilde{\ell} = \ell$.

Comparability of Information Costs

Using the Legendre transform $\tilde{\ell}$, get a pair of inequalities bracketing any information cost function $H(p)$, regardless of the convexity or subexponentiality of ℓ:

Theorem 8 *Let H be any additive information cost function determined by concave nonnegative $h = h(t)$, and put $\ell(t) = h(t)/t$ for $0 < t < 1$. For any $d \in (0,1)$, and any pdf p which is $(1-d)$-summable, we have the inequalities*

$$\tilde{\ell}\left(\sum_n p_n^2\right) \leq H(p) \leq -\frac{1}{d}\tilde{\ell}\left(\sum_n p_n^{1-d}\right),$$

where $\tilde{\ell}$ is the Legendre transform of ℓ on $(0,1)$ and of the $(-d)$-subexponential extension of ℓ on $(1,\infty)$. □

Application

For $d = \frac{1}{3}$, the previous theorem says:

$$\tilde{\ell}\left(\sum_n p_n^2\right) \leq H(p) \leq -3\tilde{\ell}\left(\sum_n p_n^{2/3}\right).$$

Suppose p, q are $1 - d = 2/3$-summable pdfs. To show that $H(p) \leq H(q)$ for all H in some class, it suffices to show that

$$-3\tilde{\ell}\left(\sum_k p_k^{\frac{2}{3}}\right) \leq \tilde{\ell}\left(\sum_k q_k^2\right),$$

for all ℓ derived from the class.

Now, Hölder's inequality implies that

$$1 = \sum_k q_k^{\frac{1}{2}} q_k^{\frac{1}{2}} = \left(\sum_k q_k^{\frac{2}{3}}\right)^{\frac{3}{4}} \left(\sum_k q_k^2\right)^{\frac{1}{4}},$$

so $\sum q_k^2 = \left(\sum_k q_k^{\frac{2}{3}}\right)^{-3}$.

Also, $h_0(t) = t^{2/3}$ defines an additive information cost function $H_0(p) = \sum_k p_k^{\frac{2}{3}} \geq 1$.

Thus $H(p) \leq H(q)$ for every information cost function H defined by $h(t) = t\ell(t)$ with

$$-3\tilde{\ell}\left(H_0(p)\right) \leq \tilde{\ell}\left(H_0(q)^{-3}\right).$$

For the class $\{h_T\}$, the function $\ell(t) = h_T(t)/t$ satisfies:

$$\ell(t) = \begin{cases} 1, & \text{if } 0 \leq t < T, \\ T/t, & \text{if } T \leq t \leq 1. \end{cases}$$

$$\tilde{\ell}(t) = \begin{cases} 1 - (1-T)t, & \text{if } 0 \leq t \leq 1 \leq 2T, \\ 1 - \frac{t}{4T}, & \text{if } 0 \leq t \leq 2T < 1, \\ \ell(t), & \text{if } 2T < t \leq 1. \end{cases}$$

The $(-d)$-subexponential extension to $(1,\infty)$ of ℓ, and its Legendre transform $\tilde{\ell}$, will be

$$\begin{aligned}
\ell(s) &= \begin{cases} -dTs^{1/d}, & \text{if } 1 \leq s \leq T^{-d}, \\ -d, & \text{if } T^{-d} < s < \infty, \end{cases} \\
-\frac{1}{d}\tilde{\ell}(s) &= \begin{cases} T + \frac{1-T}{T^{-d}-1}(s-1), & \text{if } 1 \leq s \leq T^{-d}, \\ 1, & \text{if } T^{-d} < s < \infty. \end{cases}
\end{aligned}$$

If $H_0(p) < H_0(q)$, then an open interval of values $T \in (0,1)$ satisfy $H_0(q)^{-3} \leq T \leq H_0(p)^{-3}$. But then for H defined by such h_T,

$$\begin{aligned}
-3\tilde{\ell}\,(H_0(p)) &= T + \frac{1-T}{T^{-1/3}-1}(H_0(p)-1) \\
&\leq T + (1-T) \;=\; 1 \\
&\leq T/H_0(q)^{-3} \\
&= \tilde{\ell}\left(H_0(q)^{-3}\right).
\end{aligned}$$

The result follows.

Further Reading

1 J. Aczél and Z. Daróczy. *On Measures of Information and Their Characterizations*. Academic Press, New York, 1975.

2 R. R. Coifman and M. V. Wickerhauser. Entropy based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 32:712–718, March 1992.

3 L. Fuchs. A new proof of an inequality of Hardy–Littlewood–Pólya. *Mat. Tidsskr.*, B:53–54, 1947.

4 G. H. Hardy, J. E. Littlewood, and G. Pólya. Some simple inequalities satisfied by convex functions. *Messenger of Mathematics*, 58:145–152, 1929.

5 G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, England, 1959.

6 A. S. Markus. Eigenvalues and singular values of the sum and product of linear operators. *Uspehi Mat. Nauk*, 19(4):93–123, 1964. Russian Math. Surveys 19:91–120, 1964.

7 A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications*. Number 143 in Mathematics in Science and Engineering. Academic Press, New York, 1979.

8 A. Rényi. *Wahrscheinlichkeitsrechnung*. Veb Deutscher Verlag der Wissenschaften, Berlin, 1962.