174.

Updating Hardy, Littlewood and Pólya with Linear Programming

Larry Shepp

Abstract Von Neumann's development of linear programming [8] postdates the great book on inequalities of Hardy, Littlewood, and Pólya [7]. I believe that there should be a systematic updating of [7] using the linear programming or convexity methodology that I will illustrate here with examples and that this approach will produce new insights or at least new proofs. I expect all the inequalities of Chapter 1 of [7], at least, can be proved by linear programming methods, but I am not sure of this and discuss some known inequalities that I do not know how to do with the linear programming method. At the end, I give an example of an inequality that provably *cannot* be done with convexity methods, for which I have no proof at all, just convincing evidence that it is true.

1 Proving Two Simple Inequalities with Convexity Methods

Some of the the standard inequalities that mathematicians use can be proven with convexity arguments or linear programming. Perhaps others cannot, so we might say that an inequality is "simple" if there is a convexity based proof. The Cauchy-Schwarz inequality, which may be the most famous and useful inequality ever found is simple in this sense [12], but there are so many proofs of it that it seems that almost any method will give one, so it may be that it is simple in any sense. The Schwarz inequality can be stated for a general measure space but it easily reduces to the statement that
$EX^2EY^2 \geq (EXY)^2$,

Larry Shepp
Rutgers University, Statistics Department, Hill Center, Busch Campus, 110 Frelinghuysen Road, Piscataway NJ 08854 USA, e-mail: shepp@stat.rutgers.edu

where X and Y are any r.v.'s on a common probability space, Ω. Equality holds if and only if X and Y are proportional.

To give a convexity based proof, one thinks of the probability measure, μ, on the space, $(\Omega, \mathcal{F}, P)$, on which X and Y are defined as an element of the convex set of all probability measures. Then one finds a linear functional, $L(\mu)$, of μ, which is based on a function, $f_L(x, y)$ which must be everywhere nonnegative and whose integral, $L(\mu) = \int_\Omega f_L(X, Y) d\mu$, would therefore be nonnegative. This would then give an inequality which must then be the same as the Schwarz inequality. At this stage, this is not possible because the Schwarz inequality by its very nature is *non-linear* in μ - indeed both terms, $(EXY)^2$ and EX^2EY^2 are quadratic when looked upon as functionals of μ. Instead we must *encode* the quadraticity in the statement in order to linearize the problem and we do this by going to the product probability space, $\Omega \times \Omega$, with the product measure, so that X_1, Y_1 and X_2, Y_2 are two *independent* pairs of r.v.'s on the product space each with the joint distribution of X, Y. Note that for any four numbers, x_1, y_1, x_2, y_2, the homogeneous polynomial

$$f(x_1, y_1, x_2, y_2) = x_1^2 y_2^2 + x_2^2 y_1^2 - 2x_1 y_1 x_2 y_2 = (x_1 y_2 - y_1 x_2)^2 \geq 0,$$

is indeed nonnegative. If we now substitute X_i, Y_i for $x_i, y_i, i = 1, 2$ and take expectations, using the independence of r.v.'s with different subscripts we obtain that

$$2EX^2EY^2 \geq 2(EXY)^2$$

and the proof is complete after dividing by 2. The only case of equality is when $X_1Y_2 - X_2Y_1 \equiv 0$, that is when the ratios $\frac{X_i}{Y_i}$ are constant since they are independent for $i = 1, 2$ and it is easy to see that two independent r.v.'s which are equal must each be constant. We have linearized the problem and we have encoded all the conditions we need to make the proof work. This is what one does in linear programming, and so this is a "linear programming" or "convexity" proof. This proof appeared in the paper with Ingram Olkin [9], which uses the same technique to prove a more difficult inequality due to M. Brown, among others which I now discuss. I reproduce this material here for clarity and convenience.

M. Brown's inequality [2] states that for *positive and independent* r.v.'s on a common probability space

$$\frac{E\frac{1}{X+Y}}{E\frac{1}{(X+Y)^2}} \geq \frac{E\frac{1}{X}}{E\frac{1}{X^2}} + \frac{E\frac{1}{Y}}{E\frac{1}{Y^2}}.$$

Equality holds if and only if both X and Y are constants.

The Brown inequality is proved in [9] by the same method, and the encoding of all the conditions is similar. I refer to [9] for some details but the idea of the proof is completely analogous. We first "clear of fractions" by multiplying by $EX^{-2}EY^{-2}E(X+Y)^{-2}$, and so it is equivalent to show that

$$E\frac{1}{X+Y}EX^{-2}EY^{-2} - E\frac{1}{(X+Y)^2}(EX^{-1}EY^{-2} + EX^{-2}EY^{-1} \geq 0.$$

To use the linear functional or convexity method as above for the Schwarz inequality, we again construct the product probability space on which two independent pairs of independent r.v.'s X_i, Y_i are defined. Then *if* we could show that for any four numbers x_1, y_1, x_2, y_2 the function

$f(x_1, y_1, x_2, y_2) \equiv \frac{1}{x_1+y_1}\frac{1}{x_2^2}\frac{1}{y_2^2} - \frac{1}{(x_1+y_1)^2}(x_2^{-1}y_2^{-2} + x_1^{-2}y_2^{-1})$

is everywhere nonnegative, then it would easily follow that

$Ef(X_1, Y_1, X_2, Y_2) \geq 0$

which would then prove the Brown inequality, but (alas) f takes negative values. Alternatively, *if* we could show that $f(x_1, y_1, x_2, y_2) + f(x_2, y_2, x_1, y_1)$ is everywhere nonnegative, then the same proof would give the Brown inequality because upon substituting r.v.'s X_i, Y_i for x_i, y_i we would get the desired inequality after dividing by two. Again (alas), there are numbers $x_i, y_i, i = 1, 2$ for which this form is also negative. Fortunately, we have one last chance. If we can show that the doubly mixed (symmetric in x_1, x_2 and in y_1, y_2) form

$f(x_1, y_1, x_2, y_2) + f(x_1, y_2, x_2, y_1) + f(x_2, y_1, x_1, y_2) + f(x_2, y_2, x_1, y_1) \geq 0$

for all positive values of $x_i, y_i, i = 1, 2$, then substituting X_i, Y_i for x_i, y_i, taking expecations and using the independence of X_1, X_2, Y_1, Y_2 we get the Brown inequality after dividing by 4. The last inequality is true and the proof is easy provided one does it in the right way. I refer to [9] for the details. Note in the Brown inequality we have to encode the independence and all the symmetry of the problem, but it is all quite natural.

2 An Inequality with Both a Convexity Proof and an Alternative Proof

Another example of how linear programming methodology can provide new inequalities is taken from a paper of J. Reeds and M. Win [10]. Some recent work in wireless communications engineering [3] raised the problem of determining the best constants L and U such that

$\prod_{k=1}^{n} \frac{1}{L+a_k} \leq \int_0^\infty \prod_{k=1}^{n} \frac{1}{x+a_k} m(dx) \leq \prod_{k=1}^{n} \frac{1}{U+a_k}$

hold uniformly for all values of $a_k > 0$, $1 \leq k \leq n$, where

$m(dx) = \frac{1}{\pi\sqrt{x(1-x)}}\,dx$ on $[0, 1]$.

If we can prove the following general result for a given probability measure μ, and if it holds for $\mu(dx) = m(dx)$, then we can easily find the best values of L and U.

Given a probability measure μ on $[0, \infty)$, and positive a_k, define $c(\mathbf{a}, \mu) = c(a_1, \ldots, a_n, \mu)$ to be that positive real value of c such that

$\int_0^\infty \prod_{k=1}^{n} \frac{c+a_k}{x+a_k} \mu(dx) = 1.$

Theorem 1. *For any probability measure μ, $c(\mathbf{a}, \mu)$ is monotone increasing in each a_k.*

More precisely, $c(a_1, \ldots, a_n, \mu)$ is defined by the implicit equation

$H(a_1, \ldots, a_n, \mu, c(a_1, \ldots, a_n, \mu)) = 1,$

where H is defined by

$H(a_1, \ldots, a_n, \mu, c) = \int_0^\infty \prod_{k=1}^{n} \frac{c+a_k}{x+a_k} \mu(dx).$

One can interpret $c(a_1, \ldots, a_n, \mu)$ as a *generalized mean*, $\mathcal{M}_\phi[\mu]$, in the sense of [7], Chap. VI, for a suitable function $\phi(x) = \phi(x, a_1, \ldots, a_n)$. A generalized mean wrt. a strictly monotonic function $\phi(x)$, denoted $\mathcal{M}_\phi[\mu]$, is defined by the equation

$\phi(\mathcal{M}_\phi[\mu]) = \int \phi(x)\mu(dx)$.

In our case, $\phi(x, a_1, \ldots, a_n) = \prod_{k=1}^n \frac{1}{x+a_k}$ for $x \geq 0$ gives $\mathcal{M}_\phi[\mu] = c(a_1, \ldots, a_n, \mu)$.

We prove by linear programming arguments that the desired monotonicity follows for any probability measure μ. It can alternatively be proved by applying a criterion of [7], Chap. III, which gives necessary and sufficient conditons on pairs of monotonic functions (ϕ, ψ) for $\mathcal{M}_\phi[\mu] \leq \mathcal{M}_\psi[\mu]$ for all μ. For details of the alternative proof see [10] [Appendix]. Our proof seems simpler but it is also indirect.

We reformulation the problem by first implicitly differentiating H with respect to a_i which gives

$\frac{\partial H(\mathbf{a},\mu,\mathbf{c})}{\partial a_i}|_{c=c(\mathbf{a},\mu)} + [\frac{\partial c(\mathbf{a},\mu)}{\partial a_i}]\frac{\partial H(\mathbf{a},\mu,\mathbf{c})}{\partial c} = 0.$

Clearly $\frac{\partial H}{\partial c}(\mathbf{a}, \mu, \mathbf{c}) > \mathbf{0}$, so to show that $c(\mathbf{a}, \mu)$ is increasing in a_i it suffices to show that

$H_i \equiv \frac{\partial H(\mathbf{a},\mu,\mathbf{c})}{\partial a_i}|_{c=c(\mathbf{a},\mu)} \leq 0,$

whenever $H(\mathbf{a}, \mu, \mathbf{c}) = \mathbf{1}$.

This leads to consideration of the set $\mathcal{C}(\mathbf{a}, \mathbf{c}) \subset \mathcal{R}^{\mathbf{2}}$ of possible values of the pair

$(H(\mathbf{a}, \mu, \mathbf{c}),\ \mathbf{H_i}(\mathbf{a}, \mu, \mathbf{c}))$ as μ ranges over all probability measures, for fixed c and $\mathbf{a}$. Indeed, the closure of $\mathcal{C}(\mathbf{a}, \mathbf{c})$ is the convex hull of the union of $(0,0) \in \mathcal{R}^2$ and the curve

$$x \mapsto (h(x),\ h_i(x))$$

in $\mathcal{R}^2$ traced out by $x \in [0, \infty)$, where

$$h(x) \ = \ \prod_{k=1}^{n} \frac{c + a_k}{x + a_k} \tag{1}$$

and

$$h_i(x) \ = \ \frac{\partial}{\partial a_i}\left(\prod_{k=1}^{n} \frac{c + a_k}{x + a_k}\right)\Bigg|_{c=c(\mathbf{a},\mu)} .$$

Then

$$H = \int_0^\infty h(x)\mu(dx)$$

and

$$H_i = \int_0^\infty h_i(x)\mu(dx)\,.$$

The main fact that we use about $\mathcal{C}(\mathbf{a}, \mathbf{c})$ is a linear inequality.

Lemma 1. *There exists a $\lambda > 0$ such that for all $(s,t) \in \mathcal{C}(\mathbf{a},\mathbf{c})$,*

$$s + \lambda t \leq 1 .$$

The Theorem follows directly from the Lemma since from

$$H(\mathbf{a},\mu,\mathbf{c}) + \lambda\,\mathbf{H_i}(\mathbf{a},\mu,\mathbf{c}) \leq \mathbf{1}$$

and $H(\mathbf{a},\mu,\mathbf{c}) = \mathbf{1}$ it follows that $H_i(\mathbf{a},\mu,\mathbf{c}) \leq \mathbf{0}$, as desired.

Proof. We will exhibit a $\lambda > 0$ such that for all $x \geq 0$,

$$h(x) + \lambda h_i(x) \leq 1 \tag{2}$$

and

$$h_i(x) = \left[\frac{1}{c + a_i} - \frac{1}{x + a_i}\right]\, h(x) . \tag{3}$$

Integrating (2) against μ will then finish the proof.

To see that such a λ exists, we rearrange (2) into the equivalent form

$$x + a_i + \lambda\frac{x - c}{c + a_i} \leq (x + a_i)\prod_{k=1}^{n}\frac{x + a_k}{c + a_k} . \tag{4}$$

The right hand side of (4), which we are to prove, is a polynomial function in x with positive coefficients, and hence convex on $[0,\infty)$. The left hand side is an affine function of x, agreeing with the right hand side when $x = c$. With appropriate choice of λ the left hand side's derivative matches the right hand side's derivative at $x = c$, too. For that choice of λ, (4) will hold for all $x \geq 0$.

It remains to check the positivity of the chosen λ, namely, of the solution of

$$\left.\frac{d}{dx}\right|_{x=c} x + a_i + \lambda\frac{x - c}{c + a_i} \;=\; \left.\frac{d}{dx}\right|_{x=c} (x + a_i)\prod_{k=1}^{n}\frac{x + a_k}{c + a_k} ,$$

that is, of

$$1 + \frac{\lambda}{c + a_i} = 1 + (c + a_i)\sum_{k=1}^{n}\frac{1}{c + a_k} .$$

But clearly

$$\lambda = (c + a_i)^2\sum_{k=1}^{n}\frac{1}{c + a_k} > 0 ,$$

which finishes the proof.

3 Discussion

The argument given here follows the pattern of a typical application of the "weak duality theorem" of finite dimensional linear programming. Finite dimensional linear programming deals with the problem of maximizing a linear form such as (c, x) with respect to $x \in \mathcal{R}^n$ subject to constraints of the form $\sum_{j=1}^{n} a_{ij} x_j = b_i \quad i = 1, \ldots, m x_j \geq 0 \quad j = 1, \ldots, n.$

Associated with each such problem [13], pp. 73,74, is its dual problem, that of minimizing (y, b) with respect to $y \in \mathcal{R}^m$, subject to the constraints $\sum_{i=1}^{m} y_i a_{ij} \geq c_j \quad j = 1, \ldots, n.$

This in fact is not the "standard form" for presenting primal and dual linear programming problems, but an equivalent one which matches this application more exactly.

The weak duality theorem [13], p. 58, asserts that if $x \in \mathcal{R}^n$ and $y \in \mathcal{R}^m$ satisfy the constraints of the original problem and of its dual, respectively, then $(c, x) \leq (y, b)$.

In our case, working formally and ignoring all differences between finitely many dimensions and uncountably many, consider the problem of finding a finite signed measure μ on $[0, \infty)$ which maximizes the linear form $H_i(\mu) = \int_0^\infty h_i(x)\mu(dx)$ subject to the constraints $\int_0^\infty 1\mu(dx) = 1; \int_0^\infty h(x)\mu(dx) = 1; \mu(dx) \geq 0$ for all $x \geq 0$.

The dual problem would be that of minimizing $u + v$ over $\mathcal{R}^2$, subject to the uncountably many constraints

$$u \cdot 1 + v h(x) \geq h_i(x) \quad for \quad all \quad x \geq 0\,. \tag{5}$$

The weak duality theorem would then say that if μ and (u, v) satisfied their respective constraints, then

$$\int_0^\infty h_i(x)\mu(dx) \leq u + v\,. \tag{6}$$

But the $\lambda > 0$ of the Lemma obeys (2), namely $h(x) + \lambda h_i(x) \leq 1$ for all $x \geq 0$, which means $(u, v) = (1/\lambda, -1/\lambda)$ satisfies the constraint (5). So (6) would then imply

$$H_i = \int_0^\infty h_i(x)\mu(dx) \leq 0\,,$$

which of course gives us the Theorem.

Although our proof of the Theorem and Lemma would have been perfectly comprehensible to mathematicians such as Caratheodory and Markov working in the early 1900s, the formalism of linear programming duality — which seems to have originated half a century later [13], p.87, — would not have been available to them.

4 Examples of known inequalities where there may or may not be a convexity proof

My next two examples are incomplete and suggestions for further work; I suggest trying to prove each of the Schur and FKG inequalities via convexity, which may or may not be possible.

Schur: It would be nice to see a proof of Schur's inequality via convexity. Schur's inequality (Morehead's inequality in [7]) states that if $f(x_1, \ldots, x_n)$ is permutation symmetric in its arguments $x_j \geq 0$, and differentiable, and if $\frac{\partial f(x)}{\partial x_1} \geq \frac{\partial f(x)}{\partial x_2}$ whenever $x_1 \geq x_2$, then $f(x) \geq f(y)$ whenever $y = Ax$ where A is a matrix with nonnegative entries and row sums one. An example is $f(x) = \sum_{i=1}^n x_i^2$. It's clear that the set, C, of all such f is convex and so there has to be a linear functional,

$F_{x,y}(g) = \int g(x, y, z)\mu_{x,y}(dz)$

for which the hyperplane $f(x) - f(y) \geq 0$ for all $f \in C$. Indeed the usual proof as in [7] gives such a μ, if written from this point of view, see also [9]. Is μ unique?

FKG: It would be nice to see a proof of the FKG [6], inequality via convexity. This extremely useful inequality came along after [7]. Ahlswede and Daykin [1] gave a later and sometimes more useful version [5], but I will discuss only the original version. Here Λ is a partially ordered set which is also a distributive lattice, and $\mu \geq 0$ is a function on $x \in \Lambda$ satisfying the property $\mathcal{P}$: $\mu(x \vee y)\mu(x \wedge y) \geq \mu(x)\mu(y)$, where $x \vee y$ is the unique largest element of Λ smaller than both x and y and $x \wedge y$ is the unique smallest element of Λ larger than both. If now f, g are a pair of monotinic functions on Λ, $f(x) \leq f(y)$ and $g(x) \leq g(y)$ when $x < y$, then the FKG inequality asserts that

$\sum_{x\in\Lambda} f(x)g(x) \sum_{y\in\Lambda} \mu(y) \geq \sum_{x\in\Lambda} f(x)\mu(x) \sum_{y\in\Lambda} g(y)\mu(y)$.

A new proof has been given by Siddharti Sahi [11] in this volume, but I would like to see a convexity proof. There are at least two approches to a convexity proof: one can look at the convex set of all f satisfying the given inequalities with g, μ held fixed; or one can fix f, g and try to imitate the proof of the Schwarz or Brown inequality. It is well known that property $\mathcal{P}$ is not "sharp", but how to make a better general condition on μ for the same conclusion is unclear. It's a long shot, but maybe convexity can help.

An example where we can show that there is no convexity proof.

Finally, I discuss a conjectured inequality that I will show *cannot* be proved by using convexity. I believe it to be true but I have no proof for it. This example has not been published before and I owe the formulation of the problem to J. Denny.

Let X, Y denote a pair of symmetric (about zero) independent r.v.'s with common variance σ^2, given, and with cumulative distribution functions, F, G, respectively. Define the functional $\mathcal{F}(F,G) = P(X+Y \geq 1)$, and let $\phi(\sigma)$ denote the supremum of $\mathcal{F}(F,G)$ over all such choices of F, G, so that,

$\phi(\sigma) = \sup_{X,Y} P(X+Y \geq 1)$

under the constraints that each of X and Y are symmetric about zero, each has variance σ^2, and X, Y are independent. We want to explicitly find $\phi(\sigma)$. First note that

$\mathcal{F}(F,G) = P(X+Y \geq 1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \chi(x+y \geq 1)F(dx)G(dy)$

is bi-linear in the pair F, G but is not convex. This causes the problem and it becomes even worse (quadratic) if we impose $F = G$ as is the case if $X \sim Y$ are assumed to have the same distribution. In case F and G can be chosen separately, the for each fixed F, $\mathcal{F}(F,G)$ is linear in G and for each fixed G, $\mathcal{F}(F,G)$ is linear in F. Since the class of symmetric distributions with given variance is convex, a theorem of Lester Dubins [4], says that for the extreme "points" at which the supremum of $\mathcal{F}(F,G)$ must occur, each of X and Y have at most 4 values. Thus the answer can be *numerically* determined via a search through a 4 dimensional parameter space. This leads me to believe that I know the exact formula for $\phi(\sigma)$ for every σ, but a rigorous proof that the formula is correct escapes me. For $\sigma < c$ where $c \sim .724$, I indicate next that a rigorous proof might be devised based on linear programming, but for $\sigma > c$, it can be *proven* that the linear programming argument breaks down and some other method of proof has to be found. I do not know how to move ahead to find an honest proof that the search gives the correct maximum.

If we relax the independence condition on X, Y, which is a non-linear condition, and replace it by the weaker condition that

$E[X^2 - \sigma^2|Y] \equiv 0$ and

$E[Y^2 - \sigma^2|X] \equiv 0$

which of course holds if X, Y are independent, then the problem becomes a linear programming problem (infinite dimensional, but linear). We have linearized the problem and in the wider class of X, Y pairs satisfying the last two conditions if it turns out that the maximum value is just the one obtained by the search then this in principle would give a rigorous proof that the search found the optimum. For $\sigma <\sim .724$ the upper bound seems to coincide with the value for $\phi(\sigma)$ obtained from the search (at least numerically) which would then give a rigorous proof that there is no "duality gap" between the linear and nonlinear problems. But we show that a duality gap does exist for larger σ. The best pair under the relaxed conditions on conditional expectations above are no longer independent for $\sigma > .724$. Some method other than linear programming will be required to give a rigorous proof for these values of σ.

We find convincingly that the maximum value of $\phi(\sigma) = P(X+Y \geq 1)$ taken over all pairs of independent symmetric r.v.'s X, Y each with variance $\leq \sigma^2$ strictly increases in σ for $0 \leq \sigma \leq 1$, and thereafter, i.e., for $\sigma \geq 1$, $\phi(\sigma) \equiv \frac{1}{2}$, although there is no i.i.d. pair that achieves the value $\frac{1}{2}$ for any value of σ. The range $[0,1]$ breaks up into 5 intervals, $(\sigma_i, \sigma_{i+1}), i = 0, \ldots, 4$, with $\sigma_0 = 0, \sigma_1 = \frac{1}{2}, \sigma_4 = 1$. In each interval the optimal X, Y has a different form and in the first, third, and fourth interval the optimal pair are i.i.d., but this is not so for the second and fifth intervals, which seems surprising. More precisely, we have that

I. for $0 \leq \sigma \leq \sigma_0$ an optimal X, Y pair (not necessarily unique) is *identically distributed* and each of X, Y is equal to ± 1 w.p. $\frac{\sigma^2}{2}$ and equal to zero w.p. $1-\sigma^2$. It is easy to verify that this makes $\phi(\sigma) = \sigma^2 - \frac{3}{4}\sigma^4$ for $0 \leq \sigma \leq \sigma_0$.

II. for $\sigma_0 < \sigma \leq \frac{1}{2} = \sigma_1$, surprisingly, the (apparently unique up to an interchange of X, Y) optimal X, Y pair is *not* identically distributed (except at $\sigma = \frac{1}{2}$). One of the variables, say X has the distribution $X = \pm\sigma$ w.p. $\frac{1}{2}$, while Y has the unequal distribution $Y = \pm(1-\sigma)$ w.p. $\frac{\sigma^2}{2(1-\sigma)^2}$ and $Y = 0$ w.p. $1 - (\frac{\sigma}{1-\sigma})^2$. Of course X and Y could be interchanged here so there are at least two different optimal pairs now. It is easy to check that for σ in this range this gives $\phi(\sigma) = \frac{1}{4}(\frac{\sigma}{1-\sigma})^2$. Here $\sigma_0 \sim .46$ is the root of the quartic equation

$\sigma^2 - \frac{3}{4}\sigma^4 = \frac{1}{4}(\frac{\sigma}{1-\sigma})^2$,

required to make ϕ continuous at σ_0 (smooth-fitting).

III. for the next range, $\frac{1}{2} \leq \sigma \leq \sigma_1 \sim .65$, the optimal X, Y pair are again identically distributed and

$X \sim Y = \pm\frac{3}{2}$ w.p. $\frac{1}{4}(\sigma^2 - \frac{1}{4})$ and

$X \sim Y = \pm\frac{1}{2}$ w.p. $\frac{1}{4}(\frac{9}{4} - \sigma^2)$.

This gives the value $\phi(\sigma) = \frac{23}{128} + \frac{5}{16}\sigma^2 - \frac{1}{8}\sigma^4$.

IV. for the next range, $\sigma_1 \leq \sigma \leq \sigma_2 \sim .78$, the optimal pair X, Y is again identically distributed and has the same distribution as the earlier identical pair for $0 \leq \sigma \leq \sigma_0$ and the same formula for ϕ, $\phi(\sigma) = \sigma^2 - \frac{3}{4}\sigma^4$.

V. for $\sigma_2 \leq \sigma \leq 1$, the optimal pair is not identically distributed and have the distributions

$X = \pm 1$ w.p. $\frac{1}{2}\sigma^2$, $X = 0$ w.p. $1 - \sigma^2$,

$Y = \pm 2$ w.p. $\frac{\sigma^2}{8}$, $Y = 0$ w.p. $1 - \frac{\sigma^2}{4}$.

This gives the value for this range of σ as

$\phi(\sigma) = \frac{5\sigma^2 - \sigma^4}{8}, \sigma_2 \leq \sigma \leq 1$.

The above particular choices of the pair X, Y give a lower bound on $\phi(\sigma)$. In the next section we will use linear programming, or duality, to obtain an upper bound which is tight for some σ, but alas, not for all σ. The upper bound will be a "linear programming bound" despite the fact that we are maximizing a non-linear (actually bilinear) functional:

$P(X + Y \geq 1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \chi(x + y \geq 1) F(dx) G(dy)$

which is linear in the d.f. F of X for each fixed d.f. G of Y, and is linear in G for each fixed F, but is not linear in the pair F, G since it involves products. It is even "less" linear if $X \sim Y$ is imposed as an extra condition - more on this later. It is well-known that a linear functional on a compact convex set C is maximized at an extreme point of C. Here C might be the set of pairs F, G with G fixed and F the d.f. of any symmetric X with variance σ. It's a consequence of the Dubins-Caratheodory theorem that an extreme point of C is a distribution with at most 4 points in its support which is a set of the form $\{-b,-a,a,b\}$ or $\{-a,0,a\}$, or $\{-a,a\}$ or (for $\sigma=0$, just $\{0\}$). Similarly for each fixed F the best G has at most four points in its support. It follows rigorously that the best F, G are each "four-pointers". A search of this set of distributions for each of F, G led to the conjecture above for $\phi(\sigma)$. Each of the above 4 particular choices for X, Y gives a lower bound for $\phi(\sigma)$, and the maximum for each σ of the 4 lower bounds is also a lower bound for $\phi(\sigma)$ which we call $\phi_0(\sigma)$. Note that pair I (or IV) can be used for any value of $\sigma \leq 1$ but II is allowed for $\sigma \leq \frac{1}{2}$, III for $\sigma \in [\frac{1}{2}, \frac{3}{2}]$, and V for $\sigma \in [\frac{1}{2}, 1]$.

We summarize this as follows, $\phi(\sigma) \geq \phi_0(\sigma)$, where:
$\phi_0(\sigma) = \sigma^2 - \frac{3}{4}\sigma^4, 0 \leq \sigma \leq \sigma_0,$
$\phi_0(\sigma) = \frac{1}{4}\frac{\sigma^2}{(1-\sigma)^2}, \sigma_0 \leq \sigma \leq \sigma_1 = \frac{1}{2},$
$\phi_0(\sigma) = \frac{23}{128} + \frac{5}{16}\sigma^2 - \frac{1}{8}\sigma^4, \frac{1}{2} \leq \sigma \leq \sigma_2,$
$\phi_0(\sigma) = \sigma^2 - \frac{3}{4}\sigma^4, \sigma_2 \leq \sigma \leq \sigma_3,$
$\phi_0(\sigma) = \frac{5\sigma^2-\sigma^4}{8}, \sigma_3 \leq \sigma \leq 1,$
where the values of $\sigma_i, i = 0,1,2,3$ are defined to make $\phi_0(\sigma)$ continuous, and each is the solution of an algebraic equation of degree at most four. Note also the value of $\phi_0(\sigma) \equiv \frac{1}{2}$ for $\sigma \geq 1$ and since $P(X+Y \geq 1) \leq \frac{1}{2}$ because $X+Y$ is symmetric, we see that $\phi(\sigma) \equiv \frac{1}{2}$ for $\sigma \geq 1$.
Remark: Under the additional constraint that $X \sim Y$ are equi-distributed it is impossible to make $P(X+Y \geq 1) = \frac{1}{2}$. Maximizing $P(X+Y \geq 1)$ under this additional constraint is a harder problem because this functional of the d.f. of X is non-liear (quadratic). The optimal X is not necessarily a four-pointer.

A graph of $\phi_0(\sigma)$ is shown in Figure 1.

figure=figmaxprob.eps,angle=-90

Fig. 1 The maximum value of $P(X+Y \geq 1)$ as a function of σ for independent and symmetric r.v.'s X, Y with variance σ^2.

At this point we only know that $\phi(\sigma) \geq \phi_0(\sigma)$. We wanted to show that the two functions are the same, but we now believe that this is actually false in general although it seems to hold for $\sigma <\sim .724$. In the course of conducting the search our "best" guess kept getting better especially as we also studied upper bounds given in the next section. Searches are only as good as the searcher, and it's better to have a rigorous proof that you have found the least upper bound.

Upper bounds

Even though the above problem is not linear, as remarked above, we may use a Chebyshev method, or a "duality" method to obtain upper bounds for $\phi(\sigma)$. Supppose for a given value of θ there is a function $f = f(x), 0 \leq x < \infty$

for which $|f(x)| \leq A + Bx^2$ for some fixed constants A, B, and for which for all values of $x, y \in (-\infty, \infty)$,

$f(x)(y^2 - \sigma^2) + f(y)(x^2 - \sigma^2) + \theta$

$\geq \frac{1}{4}(\chi\{x + y \geq 1\} + \chi\{x - y \geq 1\} + \chi\{-x + y \geq 1\} + \chi\{-x - y \geq 1\})$.

Then θ is an upper bound on $\phi(\sigma)$. The proof is to observe that if $X(\omega), Y(\omega)$ are defined on some probability space Ω, and X, Y are independent, symmetric, and have variance σ^2, then we may set $x = X(\omega), y = Y(\omega)$ and take expectations. By the symmetry, each of the four expectations on the right is the same, i.e. $P(X + Y \geq 1$, and so this immediately gives that for any such X, Y pair, $P(X + Y >= 1) \leq \theta$. The infimum of all such θ is called $\phi_1(\sigma)$ and we therefore have

$\phi_0(\sigma) \leq \phi(\sigma) \leq \phi_1(\sigma)$.

We can make the problem slighly simpler if we take into account the symmetry. All we really need is to find $f(x)$ for a given θ such that for all $x, y \in [0, \infty)$, rather than on the whole line, we have

$$f(x)(y^2 - \sigma^2) + f(y)(x^2 - \sigma^2) + \theta \geq \frac{1}{4}(\chi\{x + y \geq 1\} + \chi\{|x - y| \geq 1\}). \quad (7)$$

Then we can merely set $f(x) = f(-x)$ to define f for negative values of x, and we can easily verify that the first set of inequalities hold for all x, y. Note that we do not expect that f will be unique. We remark that, of course, for $\theta < \phi_0(\sigma)$ such an f cannot exist.

If for some σ, the *strict* inequality $\phi_0(\sigma) < \phi_1(\sigma)$ then there is a "duality gap". The problem of minimizing θ subject to the linear constraints above *is* a (continuous) linear programming problem in the infinitely many unknowns, $f(x), 0 \leq x < \infty$ and θ. Minimizing θ, which is a linear form in the unknowns, subject to the inequalities (7) above for all x, y is thus a linear problem. The dual problem is equivalent to maximizing $P(X' + Y' \geq 1)$ over all jointly distributed symmetric r.v.'s X', Y' which now are not necessarily independent but which satisfy $E[X^2 - \sigma^2 | Y = y] \equiv 0$ and $E[Y^2 - \sigma^2 | X = x] \equiv 0$ for all $x, y \in [0, \infty)$. To see this, observe that if we multiply in (7) by $t(x, y) \geq 0$ for x, y in a discrete finite set $\mathcal{S}$ and add we get the latter problem provided that

$\sum_{x,y} t_{x,y} = 1$,

$\sum_x t_{x,y}(x^2 - \sigma^2) \equiv 0$, and

$\sum_y t_{x,y}(y^2 - \sigma^2) \equiv 0$,

which becomes the latter problem if we interpret $t_{x,y}$ as $P(X = x, Y = y)$. This conditional expectation condition is of course weaker than independence, if X, Y are actually independent then the conditional expectations vanish because the variances are σ^2. Thus the dual version of the upper bound problem drops the independence assumption. Does this increase the value of $\phi(\sigma)$ and leave a "duality gap"? Alas there is gap for some values of σ as we will see below, but for small $\sigma <\sim .724$ there seems to be no gap.

We believe one can produce $f(\cdot, \sigma)$ for $\sigma <\sim .724$, which satisfies (7), *at least numerically* on a fine lattice x, y consisting of all multiples of a small spacing. This seems to leave little doubt that there is no duality gap for these cases. A rigorous proof that the inequalities (7) hold for the given f's for all x, y and these σ's still needs to be supplied.

For the first range, $0 \leq \sigma \leq \sigma_0$, the f, which was found by discretizing the problem to a finite set of values, $x_1, \ldots, x_n$, and then solving the linear programming problem of finding the least θ for which the inequalities (7) hold for some values $f_i = f(x_i), i = 1, \ldots, N$. We then guessed the solution to the continuous version by finding the set of i, j for which equality holds in the inequalities with $x = x_i, y = x_j$.

This led to the guess that the inequalities (7) hold with equality when $0 \leq \sigma \leq \sigma_0$ when $y = y(x) = x, x \geq \frac{1}{2}$, which indicates that, for

$\theta_0 = \phi_0(\sigma) = \sigma^2 - \frac{3}{4}\sigma^4$,

$f(x) = \frac{\frac{1}{2}-\theta_0}{2(x^2-\sigma^2)}, x \geq \frac{1}{2}$.

The linear programming solution also indicated that there is a function, $a = a(\sigma)$ for $0 \leq \sigma \leq \sigma_0$ such that for $0 \leq x \leq a$, equality holds in (7) when $y = y(x) = x + 1$. This indicates that

$f(x) = \frac{\frac{1}{2}-\theta_0-\frac{(\frac{1}{4}-\theta_0)(x^2-\sigma^2)}{((1+x)^2-\sigma^2)}}{(1+x)^2-\sigma^2}, 0 \leq x \leq a,$

and also when $a \leq x \leq \frac{1}{2}$, that equality holds when $y = 1 - x$ which, in turn, indicates that

$f(x) = (\frac{1}{4} - \theta_0)\frac{1-(x^2-\sigma^2)}{((1-x)^2-\sigma^2)}(1 - x)^2 - \sigma^2, a \leq x \leq \frac{1}{2},$

and $f(x)$ is defined for all $x \geq 0$. There is only one value of $a(\sigma)$ that makes f continuous and consistently defined. Thus for example for $\sigma = .1$ the value of $a(\sigma) \sim .1777$, found only numerically. A graph of $f(x, \sigma = .1)$ is given in Figure 2.

figure=figjackf.1.eps,angle=-90

Fig. 2 The function $f(x, \sigma = .1)$ in (7) giving the least upper bound.

I tried to use the same technique to guess a function $f(x) = f(x, \sigma)$ for σ in the second range $\sigma_0 < \sigma \leq \frac{1}{2}$ but I could not verify all the inequalities and I am not sure whether or not this case has a duality gap or not.

The linear program now gives different $y = y(x)$ where equality holds. It now indicates that for $x \geq 1 - \sigma$, equality holds when $y = x$, which indicates that with the new optimal value for θ,

$\theta_1 = \phi_0(\sigma) = \frac{1}{4}\frac{\sigma^2}{(1-\sigma)^2}$,

$f(x) = \frac{\frac{1}{2}-\theta_1}{2(x^2-\sigma^2)}, x \geq 1 - \sigma$.

For this second range of σ, $\sigma_0 \leq \sigma \leq \frac{1}{2}$, there is a threshold $b = b(\sigma) \leq \sigma$, analogous to the threshold $a(\sigma$ for the range $0 \leq \sigma \leq \sigma_0$ and now, for $0 \leq x \leq b$, equality holds when $y = 1 + x$, whereas for $b \leq x \leq \sigma$, the l.p. indicates that $y = 1 - x$. This leads to the guessed f obeying the equations

$f(x)((1 + x)^2 - \sigma^2) + f(x + 1)(x^2 - \sigma^2) + \theta_1 = \frac{1}{2}$, for $0 \leq x \leq b$

which determines $f(x)$ on this range since we know $f(x+1)$, *except that we don't yet know* b, in particular we now know $f(0)$. Next,

$f(x)(0^2-\sigma^2)+f(0)(x^2-\sigma^2)+\theta_1=0$, for $\sigma \le x \le \frac{1}{2}$

which determines $f(x)$ in this range. Next.

$f(x)((1-x)^2-\sigma^2)+f(1-x)(x^2-\sigma^2)+\theta_1=\frac{1}{4}$, for $b \le x \le \frac{1}{2}$, for $b \le x \le \frac{1}{2}$,

which determines $f(x)$ in the range $b \le x \le \sigma$ (noting that this range reflects into $[1-\sigma, 1-b]$ where f is known). There is a unique b which will make $f(x)$ continuous on $[0,\frac{1}{2}]$. Supposing b is so defined (there is an equation for b and we have found numerically, for example, for $\sigma=.48>\sigma_0$, that $b(\sigma)\sim .311784$. The rest of the range $(b,\frac{1}{2})$ reflects into $(\frac{1}{2},1-b)$ and since $1-b>1-\sigma$ this allows us to determine $f(x)$ on $(\frac{1}{2},1-\sigma)$. This makes f well-defined everywhere. A graph of $f(x,\sigma=.48)$ is given in Figure 3.

figure=figjackf.48.eps,angle=-90

Fig. 3 The function $f(x,\sigma=.48)$ in (7) giving the least upper bound.

It remains to determine $f(x,\sigma)$. It is an indication of trouble that the f in Figure 2 indicates a discontinuity near $x=.5$, and that (7) does not seem to hold as cleanly as in case I. Though I believe there is no duality gap for $\sigma <\sim .724$, the above method to find $f(x,\sigma)$ for this range is not a good one because it produces an extreme point of the set of functions f satisfying (7). It would be better to look for a smooth f or one in the "center" of the set of solutions of (7).

A duality gap seems to exist for $\sigma >\sim .724$

Since the problem of finding the smallest value of θ for which there exists a function $f=f(x)$ satisfying the inequalities of (7) for all nonnegative x,y is a linear programming problem (even finite if we restrict x,y to a discrete truncated lattice of values, it can be solved numerically for any fixed value of σ. We did this for various values of σ and found that the maximal value of θ is $\phi_0(\sigma)$ for $\sigma <$ around .7 or so, but for σ around .8 or so a strictly larger θ was found, and it was further found that the dual linear program was solved by a distribution of X,Y which satisfies

$$E[X^2-\sigma^2|Y]=E[Y^2-\sigma^2|X]\equiv 0 \tag{8}$$

but one for which X,Y fail to be independent. The actual example found for the case $\sigma=.8$ led to the following distribution:

$P(X=Y=0)=\alpha=\frac{4}{S}\frac{(1-\sigma^2)(4-\sigma^2)}{\sigma^4}$,

$P(X=\pm\frac{1}{2},Y=\pm\frac{1}{2})=\beta=\frac{1}{S}\frac{(4-\sigma^2)(1-\sigma^2)}{(\sigma^2-\frac{1}{4})^2}$,

$P(X=0,Y=\pm 1)=P(X=\pm 1,Y=0)=\gamma=\frac{1}{S}\frac{4-\sigma^2}{\sigma^2}$,

$P(X=\pm\frac{1}{2},Y=\pm 2)=P(X=\pm 2,Y=\pm\frac{1}{2}=\delta=\frac{1}{S}\frac{4-\sigma^2}{\sigma^2-\frac{1}{4}}$,

$P(X=\pm 1,Y=\pm 2)=P(X=\pm 2,Y=\pm 1)=\epsilon=\frac{1}{S}$, where S is chosen so that all these probabilities add to unity. It's easy to verify that this distribution of X, Y satisfies (8) and that

$P(X + Y \geq 1) = \beta + 4(\gamma + \delta + \epsilon)$.

We will call the right side of the last equation $\phi_1(\sigma)$; a graph is given in Figure 4.

figure=figcorr.eps,angle=-90

Fig. 4 The function $\phi_1(\sigma) = P(X + Y \geq 1)$ for the nonindependent pair satisfying (8).

We have verified numerically that $\phi_0(\sigma) > \phi_1(\sigma)$ for $\sigma <\sim .724$ but thereafter $\phi(\sigma) << \phi_1(\sigma)$. There is little doubt that this is right because we have run a linear program to maximize $P(X + Y \geq 1)$ over X, Y satisfying (8) and taking values in a large finite set S for various values of σ. We found that for σ slightly less than .724 the best X, Y pair are independent but this fails for σ slightly greater than .724. The value .724 is approximately the σ where $P(X + Y \geq 1)$ in the above example coincides with $\phi(\sigma)$ and occurs in the range (σ_2, σ_3). This of course just means that the upper bound given by linear programming is no longer tight. We still believe that for all σ,

$\phi(\sigma) = \phi_0(\sigma)$,

though we see no way to prove this. It seems that the only proof rests on a maximization of $P(X + Y \geq 1)$ over all 4 point symmetric independent r.v.'s.

More general versions of the problem.

Consider the problem of maximizing $P(X + Y \geq 1)$ under the restriction that X, Y are independent and symmetric and have variances less than or equal to σ^2. This will have the same answer as if the variance are equal to σ^2 because we can always place ϵ probability of X or Y far out and increase the variance without changing the value, so any value that can be achieved can also be achieved with variance equal to σ^2.

What if we impose the additional condition that $X \sim Y$ are equidistributed? It seems likely that the optimal X will be discrete and the number of points in its support will go to infinity with σ. If we denote by $\psi(\sigma)$ the maximum value of $P(X + Y \geq 1)$ over all r.v.'s X, Y, symmetric, identically distributed, and independent then it seems very hard to determine ψ except that when the maximum of the former problem is attained for iid X, Y then of course $\phi(\sigma) = \psi(\sigma)$. In particular this happens for $\sigma \leq \sigma_0$, and for $\frac{1}{2} \leq \sigma \leq \sigma_2$. But for $\sigma \geq 1$ it can be seen that

$\phi(\sigma) = \frac{1}{2} > \psi(\sigma), \sigma \geq 1$

because no iid pair can ever give $P(X + Y \geq 1) = \frac{1}{2}$.

Acknowledgements

I am grateful to each of my collaborators, Jack Denny, Peter Fishburn, Ingram Olkin, Jim Reeds, and Moe Win for permission to republish joint work under my own name and even moreso if they will remain my collaborators after they learn that I have written this paper without actually asking for this permission! I am also grateful to Jack Denny for posing the last problem as a general way to bound the probability that two different methods of measuring the same quantity, each with errors respectively, X, Y, modelled as symmetric and having variances σ^2, given, would differ by at least one, using normalized units, that is the probability that $|X - Y| > 1$. Because X, Y are assumed symmetric this problem reduces to the form given in the last section.

This paper is prepared for a Festschrift in honor of Peter Fishburn and the problems were chosen in the hopes that he will resolve them as he has so frequently done in the past.

The author was supported in part by the NSF grant 0203086 - Statistical Methods in Fast Functional MRI.

References

1. Ahlswede, R., Daykin, D.E.: An inequality for the weights of two families of sets, their unions and intersections. Zeit. Wahr. und verv. Gebiete **43**, 183-5 (1978)
2. Brown, M.: Exploiting the waiting time paradox: Applications of the size biasing transformations. Prob. in the Eng. and Informational Sciences **20**, 195-230
3. Conti, A., Win, M.Z., Chiani, M.: On the inverse symbol error probability for diversity reception. IEEE Trans. Commun. **51**, 753-756 (2003)
4. Dubins, L.E.: On extreme points of convex sets. J. Math. Anal. & Appl. **5**, 237-244 (1962)
5. Fishburn, P.C., Shepp, L.: The Ahlswede-Daykin Theorem. In: Althofer, et al. (eds.) Numbers, Information, and Complexity, pp. 501-516. Kluwer (2000)
6. Fortuin, C.M., Ginibre, J., Kastelyn, P.N.: Correlational inequalities for partially ordered sets. Comm. Math. Phys. **22**, 89-103 (1971)
7. Hardy, G.H., Littlewood, J.E., Pólya, G.: Inequalities. Cambridge Univ. Press (1934)
8. von Neumann, J., Morganstern, O.: Theory of Games and Economic Behavior. Princeton Univ. Press (1944)
9. Olkin, I., Shepp, L.: Several colorful inequalities. Amer. Math. Monthly **113**, 817-822 (2006)
10. Reeds, J., Shepp, L., Win, M.: Monotone parametric dependence of a generalized mean. Amer. Math. Monthly, submitted.
11. Siddhartha Sahi, this volume.
12. Steele, J.M.: The Cauchy-Schwarz Master Class, An Introduction to the Art of Mathematical Inequalities. Cambridge Univ. Press (2004)
13. Vanderbei, R.J.: Linear Programming: Foundations and Extensions, 2nd ed. Kluwer, Norwell MA (2001)