A geometric derivation and interpretation of Tchebyscheff's Inequality.

J. Conlon[1] and J. H. Dulá[2]

[1] Department of Economics and Finance, University of Mississippi, University, MS 38677

[2] Department of Management and Marketing, University of Mississippi, University, MS 38677

A geometric derivation and interpretation of Tchebyscheff's Inequality.

Tchebyscheff's Inequality is *de rigueur* in the first college course in statistics. This usually occurs by the second or third year in nearly all technically oriented undergraduate programs. This includes not just programs in mathematics, statistics, and engineering; but also business, economics, sociology, biology and many other areas. Tchebyscheff's Inequality is usually presented early in the course, soon after the concept of probability is introduced. The famous inequality makes its appearance as an interesting, or just curious, bound on the probability that a random variable will take on values in an interval as long as its mean and variance are known.

The proofs for Tchebyscheff's Inequality when they appear in the textbooks tend to be dry, boring, and uninspiring. In the more mathematically oriented courses, the instructor may attempt a proof. By and large, these proofs are perfunctory usually relying on arguments which deal with an integral of a nonnegative function over a subset of the interval of its limits of integration. Another proof makes Tchebyscheff's Inequality subordinate to Markov's Inequality when in fact it probably succeeded Tchebyscheff's result since Markov was his student.

This shabby treatment of a result which carries the name of one of the greats in mathematics means that much of its significance and elegance are lost on the student. Notably, it is not communicated that the inequality is the *best* inequality which is mathematically possible when the only information about the random variable is its mean and variance. It is the best in two senses; first it is actually optimum with respect to well-defined criteria (*i.e.*, "tightness"); and second, there is a distribution for which the bound is attained (*i.e.*, "sharpness"). Moreover, as we will show in this note, the bound has a simple interpretation and derivation, and can be understood using intuitive geometric arguments which permit understanding some of its proposed extensions and generalizations.

Let's formalize the notation. The random variable is X. It is one-dimensional and can either be continuous or discrete or a combination of both. Its mean is μ_X which is finite and its variance is σ_X^2 which is strictly positive. $\mathrm{E}[\cdot]$ is the expectation operator; thus the nth moment of the random variable is $\mathrm{E}[\mathrm{X}^n]$. The distribution of X is F and $\mathrm{P}(\mathrm{X} \in (a, b))$ is the probability that the random variable is in the interval (a, b). If an interval includes its endpoint then a square bracket is used; otherwise we use a parenthesis. Finally, recall that $\int_{-\infty}^{\infty} x dF(x) = \mu_X = \mathrm{E}[\mathrm{X}]$ and that $\sigma_X^2 = \mathrm{E}[\mathrm{X}^2] - \mu_X^2$.

Tchebyscheff's Inequality.

Tchebyscheff's Inequality states the following. The probability that the value of a random variable X is strictly within a distance k on either side of its mean is, at least, $1 - \frac{\sigma_X^2}{k^2}$. Notice that this result is only interesting if $k \geq \sigma_X$ since otherwise the bound is negative; something upon which we can always improve with a lower bound of zero.

The occurrence of the variance in the expression makes this a "second-order" bound. Second-order bounds such as Tchebyscheff's are called "maximin" when they represent the highest of all lower bounds guaranteed to be less than or equal to all possible values for a probability or expectation. The bounds also fall into a category known as "distribution-free" for obvious reasons. The term "semi-parametric" also applies to this bound because only a few parameters from the distribution are specified.

Before we can present the geometric arguments to derive this bound we need to review some basic results. The first is that if $g \geq f$ (i.e., g *dominates* f) then the expectation of g is greater than or equal to the expectation of f; that is, $\mathrm{E}[g] \geq \mathrm{E}[f]$. This is just the monotonicity property of integrals. The second basic result regards a standard trick in probability which consists of expressing a probability expression in terms of an expectation. This is done as follows:

$$\mathrm{P}(\mathrm{X} \in (a, b)) = \mathrm{E}[g]$$

where

$$g(x) = \begin{cases} 1 & \text{if } x \in (a, b); \\ 0 & \text{otherwise.} \end{cases}$$

To demonstrate Tchebyscheff's Inequality let us define g as follows:

$$g(x) = \begin{cases} 1 & \text{if } \mu_X - k < x < \mu_X + k; \\ 0 & \text{otherwise.} \end{cases}$$

Then, a lower bound on $\mathrm{E}[g]$ results from the expectation of a function dominated by g everywhere. This technique for generating bounds for probabilities and expectations is quite common. As we will see, it is just our ticket.

The first question is: what function should g dominate? Although any dominated (i.e., *feasible*) function works, we must remember that whatever we pick we are going to have to find its expected value. This means integrating with respect to a distribution we may not know. The choice is greatly simplified if we limit our selection to quadratic functions. This is not a casual choice; there is a lot of forethought in picking a parabola for domination by g. The main reason this is an obvious choice is that we can calculate the expectation of a parabola with respect to a given distribution without knowing the distribution! All we need is knowledge of the distribution's mean and variance. To see this recall that any one-dimensional quadratic can be expressed as

$$p(x) = a_0 + a_1 x + a_2 x^2$$

and therefore

$$\begin{aligned} \mathrm{E}[p] &= \int_{-\infty}^{-\infty} (a_0 + a_1 x + a_2 x^2) dF(x) \\ &= a_0 \underbrace{\int_{-\infty}^{\infty} dF(x)}_{=1} + a_1 \underbrace{\int_{-\infty}^{\infty} x dF(x)}_{=\mathrm{E[X]}} + a_2 \underbrace{\int_{a}^{b} x^2 dF(x)}_{=\mathrm{E[X^2]}} \\ &= a_0 + a_1 \mu_X + a_2(\sigma_X^2 + \mu_X^2) \end{aligned}$$

which means that if we know the mean μ_X and variance σ_X^2 of the distribution we know $\mathrm{E}[p]$ for any p defined by values for a_0, a_1, and a_2.

Three points along with three values are sufficient to define a unique parabola as long as the paired coordinates do not trace a straight line in the Eucledian plane. This means that the following three paired coordinates $(\mu_X - k, 0)$; $(\mu_X, 1)$, and $(\mu_X + k, 0)$ define a unique parabola. This parabola intersects the graph of g at the top of the center of the box and at each of the two lower corners. It is easy to see that this parabola is always below g and actually intersects it at the three points.

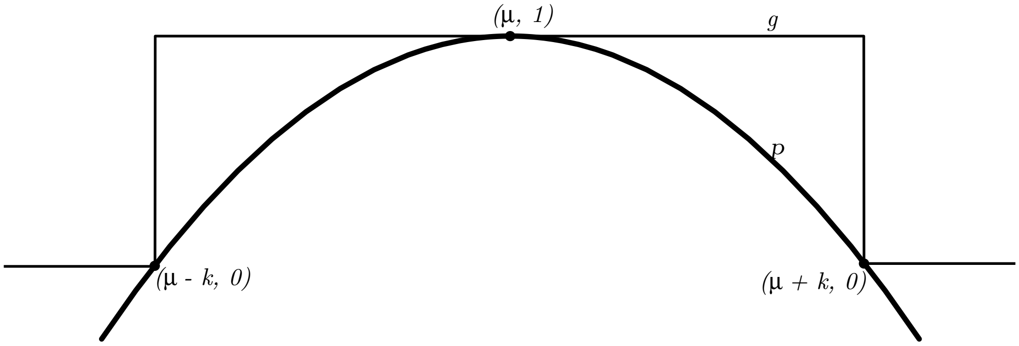

Figure 1.
The parabola p dominates the index function g and intersects it at the three points: $\mu_X - k, \mu$, and $\mu_X + k$.

The values for the three parameters, a_0, a_1, and a_2, which define the parabola are found by solving the following system of linear equations:

$$\begin{aligned} a_0 + (\mu_X - k)a_1 + (\mu_X - k)^2 a_2 &= 0 \\ a_0 + \qquad \mu_X a_1 + \qquad \mu_X^2 a_2 &= 1 \\ a_0 + (\mu_X + k)a_1 + (\mu_X + k)^2 a_2 &= 0 \end{aligned}$$

By applying Cramer's Rule to solve for the values of a_0, a_1 and a_2 we get:

$$a_0 = -\frac{\mu_X^2 - k^2}{k^2}; \quad a_1 = \frac{2\mu_X}{k^2}; \quad a_2 = -\frac{1}{k^2}.$$

This means that the expectation of the resultant parabola is

$$\mathrm{E}[p] = -\frac{\mu_X^2 - k^2}{k^2} + \frac{2\mu_X^2}{k^2} - \frac{\sigma_X^2 + \mu_X^2}{k^2} = 1 - \frac{\sigma_X^2}{k^2};$$

which is the lower bound we seek for $\mathrm{P}(\mathrm{X} \in (\mu_X - k, \mu_X + k))$.

Let's consider the following discrete random variable:

Z	$\mu_X - k$	μ_X	$\mu_X + k$
$\mathrm{P}(\mathrm{Z} = z)$	$\frac{\sigma_X^2}{2k^2}$	$1 - \frac{\sigma_X^2}{k^2}$	$\frac{\sigma_X^2}{2k^2}$

The random variable Z is well-defined only if $k \geq \sigma_X$ (remember that this is the case where Tchebyscheff's Inequality is interesting). If we calculate the mean and variance of Z we discover that these are the same as for X. Moreover, the probability that Z is in the interval $(\mu - k\sigma, \mu + k\sigma)$ is $1 - \frac{\sigma_X^2}{k^2}$; the actual value for the lower bound given by Tchebyscheff's Inequality. Therefore, there exists a random variable for which Tchebyscheff's lower bound is attained as an equality!

The fact that the lower bound predicted by Tchebyscheff's Inequality can be attained as the exact value for a probability by some random variable with the same mean and variance means the the bound is *sharp*. Sharpness in a bound in probability is good since it implies another desirable property: *tightness*. Tightness means that the bound is the best according to some criterion; sharpness implies tightness since if the bound was sharp but not the best then a better one would be theoretically possible but we would have a counterexample to it. Tchebyscheff's Inequality is therefore also tight by the criterion that it is the best lower bound from among all random variables with the same mean and variance.

Note how the distribution for Z has its mass precisely at the three points where the parabola p touches g. By allowing the three points of contact between p and g to determine the mass points of the distribution of Z, the corresponding probabilities are uniquely specified if they have to satisfy the three conditions: *i*) add up to 1, of course; *ii*) the mean must equal μ_X; and *iii*) the variance must equal σ_X^2. In general, an attempt to define a distribution based on the contact points between a feasible parabola and g is meaningful only when there are at least two such points; otherwise the variance is zero. However, there cannot be more than three contact points between g and a feasible parabola and there is only one way this can happen. Anyway, if we were to proceed in the general case to try to solve for "probabilities" such that the mean is μ_X and the variance σ_X^2 we discover that the "probability" weights are nonnegative and add up to 1 only in the case when the two functions touch at three points and $k \geq \sigma_X^2$. Moreover, the expectation of the parabola equals the "expectation" of any two or three point distribution using the, possibly spurious, probabilities since the parabola equals g at the contact points. If weights are found that are nonnegative and add up to 1 then we have a proper discrete distribution with g having the same expectation as the parabola which defines the distribution. This yields a random variable for which the lower bound is actually attained. It is sufficient for us to produce such a distribution for proof that a lower bound is sharp. The mass points of such a distribution are always defined by the points of contact between a feasible parabola and g.

Extending Tchebyscheff's Inequality.

Now that we understand how to interpret and derive Tchebyscheff's Inequality geometrically, we may try our hand at extending it. An obvious extension of what we have seen so far is to try to find a lower bound over intervals which are not symmetric around the mean; that is, a lower bound for the probability that the random variable X is within the general interval (a, b). This extension can be found in Godwin [1955] p. 928 where it is attributed to Selberg [1940].

To simplify the presentation define Y to be the simple linear transformation of X of the form: $\mathrm{Y} = \alpha\mathrm{X} + \beta$, where

$$\alpha = \frac{2k}{b-a} \qquad \text{and} \qquad \beta = -\frac{(b+a)k}{b-a};$$

this way $\mathrm{P}(\mathrm{X} \in (a, b)) = \mathrm{P}(|\mathrm{Y}| < k)$. Denote μ_Y and σ_Y^2 the mean and variance of the random variable Y, respectively, which can be calculated directly given μ_X and σ_Y^2 and the parameters of the transformation as follows: $\mu_Y = \alpha\mu_X + \beta$, and $\sigma_Y^2 = \alpha^2\sigma_X^2$. The function g becomes the inverted "box" with corners at $(-k, 0)$ and $(k, 0)$. We know we can obtain a lower bound to $\mathrm{P}(|\mathrm{Y}| < k)$

immediately by constructing a parabola dominated by g as we did in the case of the symmetric interval around the mean in the previous section. Although the resulting lower bound based on the expectation of this parabola is valid, we will see it is no longer sharp in every situation. What we require is more flexibility in the design of the dominated parabola.

As long as the parabola is feasible, the above technique will generate a valid lower bound. What we want to do is construct a parabola which will yield a lower bound such that the distribution constructed from its contact points with g actually attains it. However, we can do this in two stages: first, conjecture the appropriate form of the parabola, and second, check sharpness by constructing the appropriate discrete distribution. Since we will be demonstrating sharpness in the second stage, we only have to worry about guessing well in the first stage. That is, since sharpness is relatively easy to check in the second stage, we can be somewhat relaxed in our initial first stage construction.

Let us begin by parametrizing the feasible parabola based on where its apex touches g inside $(-k, k)$, where this may be some other point besides the center. Let us allow the parabola to also touch g at one of the two end points; whichever is closest to its apex. Let us focus on the case where the contact point γ between the parabola and the index function g occur somewhere to the right of $y = 0$. (The case for $\gamma < 0$ follows from symmetric arguments). Set another contact point at $y = k$ since this is the closest of the two end-points. Notice that we cannot have a third contact point at $y = -k$ since the parabola will then be infeasible. At this point we could use Cramer's rule to find the parameters which define the parabola (although we would need a third condition for uniqueness; namely that $p'(\gamma) = 0$). Instead, note that

$$p(y) = 1 - a(y - \gamma)^2$$

is the general form of the parabola we have defined and the condition $p(k) = 1$ means $a = 1/(k-\gamma)^2$ giving us the final form:

$$p(y) = 1 - \frac{(y - \gamma)^2}{(k - \gamma)^2};$$

John: "What about μ_Y? It's obvious that μ_Y must be positive. Let's go ahead and state it now."

José: "That condition is implied by $\gamma > 0$. It comes out later."

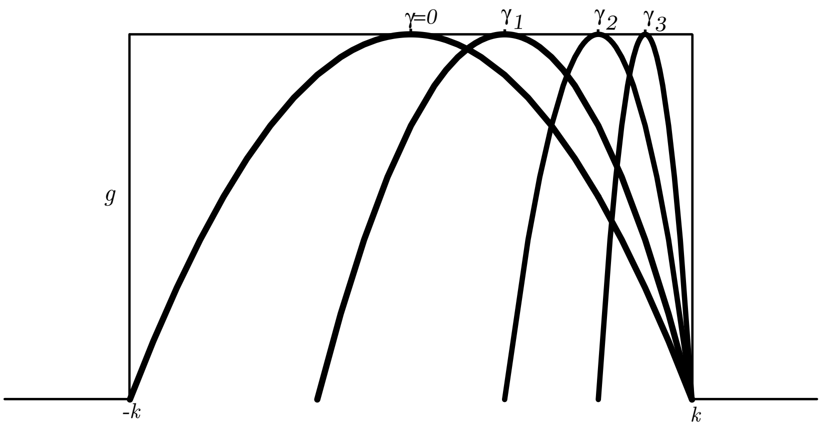

Figure 2.
Depictions of different parabolas dominated by the index function g.
As γ increases away from the midpoint, the parabolas fixed at γ and at the right endpoint, k, become "narrower."

where γ is the point of contact inside $[0, k)$. Figure 2 shows the family of parabolas parametrized by γ.

This parabola generates the following lower bound as a function of γ:

$$\mathrm{P}(|\mathrm{Y}| < k) \geq \mathrm{E}[p] = 1 - \frac{(\mu_Y - \gamma)^2 + \sigma_Y^2}{(k-\gamma)^2}; \quad 0 \leq \gamma < k. \qquad (\mathbf{LB2})$$

So far, γ is an arbitrary parameter somewhere in $[0, k)$. Any value for γ in this interval provides a valid lower bound. An interesting question is: for what value of γ do we obtain the greatest lower bound? For this we need to maximize the expression for the lower bound in **LB2** with respect to $\gamma^{\dagger}$. When we do, we obtain:

$$\gamma^* = \mu_Y - \frac{\sigma_Y^2}{(k - \mu_Y)};$$

provided it is greater than or equal to 0 (and less than k, of course). The condition $0 \leq \gamma^* < k$ generates the following inequality on μ_Y and σ_Y:

$$0 \leq \mu_Y - \frac{\sigma_Y^2}{k - \mu_Y} < k;$$

which implies that $0 < \mu_Y < k$. This can be rewritten as a condition on σ_Y^2 as follows

$$\mu_Y(k - \mu_Y) \geq \sigma_Y^2 > \underbrace{(\mu_Y - k)^2}_{<0}.$$

$\dagger$ **John**: "By the way, we don't have to check the second order conditions or anything here because later, when we check for sharpness, this will indirectly prove that γ^* gives a global max in the appropriate ranges of μ_Y and σ_Y^2."

Since $\sigma_Y^2 \geq 0$ trivially, this is simply equivalent to σ_Y^2 bounded above by $\mu_Y(k - \mu_Y)$.

Plugging the value for γ^* in the expression for the lower bound **LB2** we obtain the more specific lower bound

$$\mathrm{P}(|\mathrm{Y}| < k) \geq \frac{(k-\mu_Y)^2}{(k-\mu_Y)^2 + \sigma_Y^2}, \quad \text{for} \quad \sigma_Y^2 \leq \mu_Y(k - \mu_Y). \tag{$\mathbf{LB2}^*$}$$

For the case when $\sigma_Y^2 > \mu_Y(k - \mu_Y)$, γ^* lands to the left of the origin generating an infeasible parabola since it is anchored at the right endpoint, $y = k$. In this case we may be satisfied with the bound generated by the feasible parabola when $\gamma = 0$. Notice this case also includes $\mu_Y = 0$. Now, the lower bound we obtain based on this "symmetric" parabola is positive only when $\sigma_Y^2 + \mu_Y^2 < k^2$; otherwise, zero is a better bound (which occurs when $\mu_Y \geq 1$). This suggests the following combination of lower bounds when $\mu_Y \geq 0$:

$$\mathrm{P}(|\mathrm{Y}| < k) \geq \begin{cases} \frac{(k-\mu_Y)^2}{(k-\mu_Y)^2+\sigma_Y^2} & \text{if } \sigma_Y^2 \leq \mu_Y(k-\mu_Y) \text{ (Region 1);} \\ 1 - \frac{\sigma_Y^2+\mu_Y^2}{k^2} & \text{if } \mu_Y(k-\mu_Y) < \sigma_Y^2 < k^2 - \mu_Y^2 \text{ (Region 2);} \\ 0 & \text{if } k^2 - \mu_Y^2 \leq \sigma_Y^2 \text{ (Region 3).} \end{cases}$$

We will demonstrate that this is the best possible lower bound; that is, the bound is tight. We do this by showing that the bound is, in fact, sharp. For each of the three regions which partition the range of values for σ_Y^2, consider the following three discrete random variables[‡] and their corresponding distributions:

1. Region 1:

$\sigma_Y^2 \leq (k - \mu_Y)(\mu_Y)$

Z_1	$\mu_Y - \frac{\sigma_Y^2}{k-\mu_Y}$	k
$\mathrm{P}(\mathrm{Z}_1 = y)$	$\frac{(k-\mu_Y)^2}{(k-\mu_Y)^2+\sigma_Y^2}$	$\frac{\sigma_Y^2}{(k-\mu_Y)^2+\sigma_Y^2}$

$$\mathrm{E}[\mathrm{Z}_1] = \mu_Y, \ \mathrm{E}[\mathrm{Z}_1^2] = \sigma_Y^2 + \mu_Y^2$$
$$\mathrm{P}(|\mathrm{Z}_1| < k) = \frac{(k-\mu_Y)^2}{(k-\mu_Y)^2+\sigma_Y^2} = \mathrm{P}(\mathrm{Z}_1 = \mu_Y - \frac{\sigma_Y^2}{k-\mu_Y})$$

[‡] **José**: "Of course, we know where the mass should be: precisely at the contact points between g and the dominated parabola!"

2. Region 2:

$$(k-\mu_Y)(\mu_Y) < \sigma_Y^2 < k^2-\mu_Y^2$$

Z_2	$-k$	0	k
$P(Z_2 = y)$	$\frac{\sigma_Y^2+\mu_Y^2-k\mu_Y}{2k^2}$	$\frac{k^2-\sigma_Y^2-\mu_Y^2}{k^2}$	$\frac{\sigma_Y^2+\mu_Y^2+k\mu_Y}{2k^2}$

$$E[Z_2] = \mu_Y,\ E[Z_2^2] = \sigma_Y^2 + \mu_Y^2$$

$$P(|Z_2| < k) = 1 - \frac{\sigma_Y^2+\mu_Y^2}{k^2} = P(Z_2 = 0)$$

3. Region 3:

$$k^2-\mu_Y^2 \leq \sigma_Y^2$$

Z_3	$\sqrt{\mu_Y^2+\sigma_Y^2}$	$-\sqrt{\mu_Y^2+\sigma_Y^2}$
$P(Z_3 = y)$	$\frac{1}{2}(1+\frac{\mu_Y}{\sqrt{\mu_Y^2+\sigma_Y^2}})$	$\frac{1}{2}(1-\frac{\mu_Y}{\sqrt{\mu_Y^2+\sigma_Y^2}})$

$$E[Z_3] = \mu_Y,\ E[Z_3^2] = \sigma_Y^2 + \mu_Y^2$$

$$P(|Z_3| < k) = 0$$

At the base of each of the three tables above there is the expectation and variance of the corresponding random variable as well as the probability that it is inside the interval $(-k, k)$. In every case the expected value and variance are μ_Y and σ_Y^2; i.e., the same as Y. Also, in all three cases the probability that the random variable is in the interval $(-k, k)$ actually equals its theoretical lower bound. Therefore, in the three regions for which the values for σ_Y can be partitioned, the lower bounds we have calculated here are actually sharp.

The sharp lower bounds we have derived here can be interpreted as follows. When σ_Y^2 is in Region 1 it means that the distribution is relatively "compressed" around the mean. In this case, a "skinny" parabola, one that does not span all of the box but rather supports it at only two points, yields the best lower bound. As the dispersion of the mass of the distribution increases (σ_Y^2 enters Region 2) the parabola becomes wider reaching its widest when it supports g at the center and the two end points: $-k$ and k. When the value for the variance enters the third region, the excessive dispersion of the mass of the distribution destroys the possibility of a nontrivial lower bound.

Let us illustrate the results above with a small example. Suppose X is a random variable with mean $\mu_X = 15/2$; we will calculate lower bounds to $P(X \in (3, 9))$ for three values of σ_X^2. We begin by applying the linear transformation to X so that this probability becomes $P(|Y| < 1)$; that is, set $Y = \frac{1}{3}X - 2$. This means $\mu_Y = 1/2$ and $\sigma_Y^2 = \frac{1}{9}\sigma_X^2$. The next table includes the lower bounds based on Tchebyscheff's Inequality we have calculated above:

Region:	1	2	3
$\sigma_Y^2 =$	0.1	0.5	0.9
“Symmetric” parabola	0.65	$\underline{0.25}$	-0.15
Bound using LB2*	$\underline{0.7143}$	.333	0.2174
Trivial Bound	0	0	$\underline{0}$
Lower Bound:	0.7143	0.25	0

Note that bounds using LB2* yield the greatest value in each of the three regions in the table above. However, it is wrong to take these values as the best lower bound in every case. Expression LB2* yields a valid lower bound only in Region 1 and this bound is sharp as demonstrated by the existence of random variable Z_1 we constructed above. In Regions 2 and 3, the values obtained from LB2* correspond to the expectations of parabolas which are infeasible in the sense that they are not dominated by g everywhere as we can see in Figure 3.

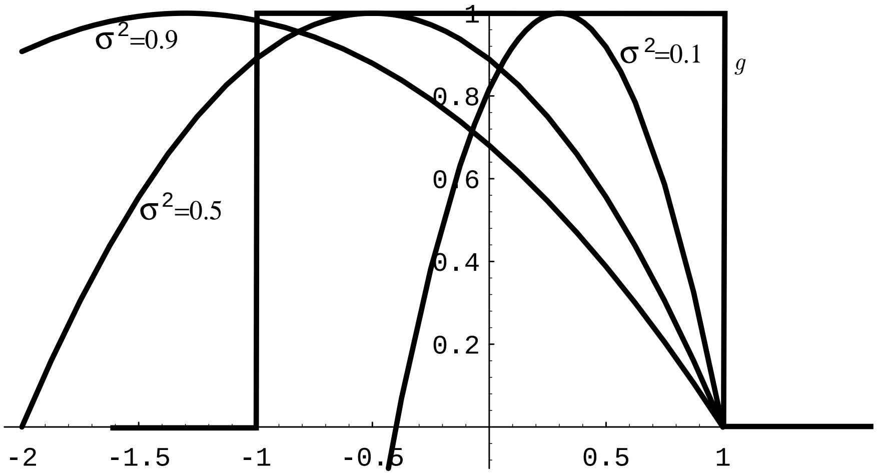

Figure 3.
The parabolas which generate the lower bounds calculated using expression LB2* when σ_Y^2 is 0.1, 0.5, and 0.9.
Note that for $\sigma_Y^2 = 0.5$ and $\sigma_Y^2 = 0.9$ the parabolas are not totally dominated by g.

The distributions we constructed to demonstrate sharpness of the bounds provide the ultimate argument as to why values from expression LB2* are not valid as lower bounds in Regions 2 & 3. In these two regions the random variables Z_2 and Z_3 we presented above actually attain the theoretical lower bounds for the probability that any random variable with mean μ_Y and variance

σ_Y^2 lies inside the interval $(-k, k)$. In both regions, these values are less than what we get from expression LB2* (in fact they are 0.25 and 0, respectively). The existence of these distributions means that the values obtained from LB2* in these two regions cannot be lower bound in these regions.

Conclusions.

Tchebyscheff's Inequality is a lower bound on the probability that any random variable with a given mean and variance lies within a specified distance on either side of the mean. We have derived this result geometrically based on the domination of a concave parabola by an index function. Besides its obvious intuitive appeal, this derivation and interpretation provides the tools to understand why Tchebyscheff's Inequality is a *sharp* bound; that is, it is the best lower bound possible when using only the mean and variance of the distribution. This geometric construction also permits an obvious extension of the traditional result. This extensions is a lower bound on the probability that the random variable is in any given interval (a, b). The result is in the form of three expressions for three possible intervals which partition the range for the variance. This geometric approach permits the demonstration that the more general lower bound is also sharp.

Other interesting extensions and generalizations which we believe the reader is now well equipped to handle include:

- Higher order lower bounds for a nonnegative random variable. This can be derived by dominating a polynomial of degree greater than or equal to 3 with an index function with corners at the origin and at k. The obvious polynomial is $p(y) = 1 - y^n/k^n$ which supports g at the origin where the polynomial has zero slope and at k, the corner of the upside-down "box" which is the graph of g. Such a lower bound requires knowledge of only the nth moment of the distribution.
- Lower bounds based on the domination of a Λ-shaped function (i.e., an inverted "V-shaped" function) with the peak somewhere inside the upside-down "box" in the graph of g. Instead of second order information, this new bound requires knowledge of the conditional expectation of two intervals; one from $-\infty$ to where the peak of the Λ-function is and the other on the other side. The bound can be greatly simplified if the index function is symmetric around the mean with the apex there as well and the distribution assumed to be symmetric (the denominators of the conditional means would be 1/2 each).

In the ideas for lower bounds in the two items above, deriving an expression for the bound is one of the challenges. However, the task will not be complete until sharpness of the bound is established.

References.

Ash, R.B., *Real Analysis and Probability*, Academic Press, New York, 1972.

Feller, W., *An Introduction to Probability Theory and its Applications*, John Wiley & Sons, Inc., New York, 1968.

Godwin, H.J., "On generalizations of Tchebychef's Inequality," *American Statistical Association Journal*, Vol. 50, September 1955, pp. 923-945.

Kemperman, J.H.B., "The general moment problem, a geometric approach," *The Annals of Mathematical Statistics*, Vol. 39, 1968, pp. 93-122.

Selberg, H.L., "Zwei Ungleichungen zur Ergänzung des Tchebycheffschen Lemmas," *Skandinavisk Aktuarietidskrift*, Vol. 23, 1940, pp. 121-125.