Averaging, Diversification, and the Median

Gilbert W. Bassett Jr.

June 2008

Abstract: Consider the estimation of a location parameter or the construction of a portfolio when uncertainty about a data or return generating process includes the possibility of fat-tailed distributions. To fix ideas, let $X_1, \ldots, X_n$ be independent and identically distributed with symmetric cdf F. Because of the uncertainty we want a location estimator to be "good" under minimal conditions on F. Let a good estimator be defined as one that is more concentrated around the unknown location than the naïve estimator that uses only a single observation. Given this definition, one question is whether the sample average $\overline{X}$ qualifies as a good, that is, better than a single observation for any F. While $\overline{X}$ may not be best, intuition suggests that averaging could never be worse than the naïve estimate based on a single observation. This intuition is closely related to the desirability of diversification; a risk averse investor should never put all eggs in one basket. Averaging data and diversification are related because a portfolio's return is the (weighted) *average* of its constituents. As is known however the intuition is wrong: a single observation can be better than $\overline{X}$, and a risk averse investor can prefer a single asset to any well diversified portfolio. One objective of the paper is to illustrate how fat-tailed distributions turn standard intuition about averaging and diversification upside down. The main objective is to indicate a positive result; namely, the median is better than a single observation, guaranteed. This is true when F is symmetric (in which case the sample average and median are estimates of the same location parameter). It is also true for the median parameter in general. As applied to diversification the result suggests a way to reformulate the conventional wisdom so that a risk averse investor should never put all their eggs in one basket.

Keywords: Averaging, Diversification, Fat Tailed Distributions, Median

Acknowledgements: Thanks to Qizhi Chen, Victor Chernozhukov, Roger Koenker, Dibyen Majumdar, John Miller, Joe Persky, Pat Rocco, and seminar participants at FERM 2008.

Department of Finance
University of Illinois at Chicago
601 South Morgan (MC 168)
Chicago, Illinois 60607-7121
(312)996-5777 FAX (312)413-7948
Email: gib@uic.edu

1. **Introduction**

Is the process of combining data via averaging always better than not combining and using a single observation? To fix ideas let $X_1,\ldots,X_n$ be independent and identically distributed with cdf F. Assume F is symmetric about an unknown μ, but is otherwise arbitrary. Let $\overline{X}$ denote the sample average and X^* the estimator based on a single randomly selected observation.[1] Intuition suggests that while $\overline{X}$ may not be the best (among all the ways of combining data), it must be better than only one observation.

Define a good estimator as one that is better than the simple benchmark estimator X^*. We know the variance of $\overline{X}$, $Var(\overline{X}) = \sigma_F^2 / n$, is less than $Var(X^*) = \sigma_F^2$ where σ_F^2 denotes the variance of X. So, the average is a good thing, at least when σ_F^2 is finite. This standard comparison however leaves open what happens when $\sigma_F^2 = \infty$ (since then the comparison of $Var(X^*) = \infty$ and $Var(\overline{X}) = \infty$ is not informative). When $\sigma_F^2 = \infty$ a dispersion criterion is also needed for comparing estimators in case moments are not absolutely integrable. With such a general criterion it is conceivable that averaging is a good principle and the problem is with the variance criterion that cannot resolve differences in estimators unless the variance of F is finite.

The answer however is known. The problem is not with the variance as a measure of dispersion, but with the average: it is not an inherently good way to combine data. Depending on the tails of F the sample mean as an estimator for μ can be no better and even much worse than the naïve X*. The best known example is the Cauchy distribution where the distribution of $\overline{X}$ and X^* are identical so that averaging is not an improvement; see e.g. Feller (1971, p.51). But the Cauchy is not the only such example. With fat enough tails the average can not only be worse than X*, but worse by an ever increasing amount as n increases.

---

[1] $X^* = X_{i^*}$ where i^* =i with probability n^{-1}, $i=1,\ldots,n$. In the iid case X^* has the same distribution as say, X_1. We use X^* as a prelude to consideration of non iid situations.

Hence, averaging is not an inherently good way to combine data. $\bar{X}$ being better than a single observation depends on conditions. Averaging can be good, even very good, relative to X^* for some F. But for other F, averaging is bad, potentially very bad.

Many people's intuition however is that averaging, if not optimal, must always be better than a single observation. Many find it hard to imagine how it could be otherwise. Without intuition about what could go wrong the conditions under which the average is inferior to X^* may be thought of no practical consequence.

Closely related to averaging is the idea of diversification. That is, let $X_1, \ldots, X_n$ now denote the annual return of n assets. Suppose for simplicity that returns are independent and follow the symmetric distribution F. Let P^A denote the return of \$1 invested in an equally weighted portfolio of each asset, so, $P^A = n^{-1}\sum_{i=1}^{n} X_i$. The diversified return is just the average of the constituent returns (the unequally weighted case will be considered later). Let P^* denote the return to a dartboard portfolio in which \$1 is invested in a randomly selected single asset. Should a risk averse investor put all their eggs in the one asset basket and earn P^*, or invest in the nicely diversified portfolio and earn P^A. Is diversification an inherently good principle?

The conventional wisdom that diversification is inherently good is perhaps even more ingrained than the intuition for averaging.[2] But the answer is the same; diversification is not a good principle. In spite of the conventional wisdom, whether or not you should put all your eggs in one basket has conditions.

The practical importance of the conditions that make diversification and averaging bad are controversial. Classical statistics identifies the mathematical boundary that separates the regions where averaging does and does not work. The cases where averaging fails are often either assumed away, or hoped to be so extreme as to be empirically implausible. The process of averaging and diversification then become, for practical purposes, good enough. Rejecting fat-tailed distributions based purely on mathematical considerations, of course, is no more justified than rejecting the Gaussian

---

[2] Diversification is axiomatic in the subadditivity requirement that defines coherent risk measures; see Artzner et al. (1997). That *conditions* underlie coherence is evident in the conditional value at risk (CVAR), which is the expected value of losses below a threshold. If the tails of the loss function are fat enough CVAR will not be able to distinguish between alternative infinite-valued CVARs.

on account of its unbounded range. Practical considerations underlie the rationale for the Gaussian in spite of its infinite range and practical considerations have to be the basis for the consideration of fat-tailed distributions.

The evidence on fat tails, especially in finance, goes back to Mandelbrot (1963) and Fama (1965). The recent book by Mandelbrot and Hudson (2004) documents a large number of more recent studies that find evidence for heavy tailed distributions. In a series of papers, Ibragimov (2006, 2005) and Ibragimov and Waldan(2007) cite the empirical evidence for fat tails and show that standard results on diversification are turned upside down in fat tailed environments; also see the recent book by Taleb(2007) that focuses on the implications of fat tails. Additional evidence as well as implications for diversification are discussed in Embrechts et al. (2007) and Neslehova, Embrechts, and Chavez-Demoulin (2006), where the emphasis is on the practical applications of fat-tailed distributions and the role they play in risk and insurance.

The most compelling evidence for fat tails emphasized in the recent studies is the distinctive data patterns they exhibit. Financial data generated by a thin tailed distribution will show a 50-50 pattern in which each day tends to contribute equally to a portfolio's overall annual return. The pattern for fat tails in contrast follows an 80-20 (or worse) rule with 80% (or more) of annual return accounted for by only 20% (or fewer) of the days. Such patterns are quite common and indicate fat tails in the underlying returns. Intuition for this feature of fat tails is provided in section 2 below.

Since averaging is not an inherently good way to combine data, one wonders if there exists any foolproof procedure guaranteed better than a single observation no matter what. Such a procedure would provide a way to combine observations or assets that would not require any conditions or assumptions for beating the naïve X*. An affirmative answer is given in Section 3 where the median is shown to be inherently good. While the proof relies on known results, the statement of the result seems new.

It is important to emphasize the result in Section 3 is not due to the sample mean being an estimate of the population mean and the sample median being an estimate of the potentially different population median. In the case of symmetry that has been considered so far (so that the population mean and median are the same μ): (i) the sample mean is not good for μ because there are symmetric F such that the one observation estimator is better than $\bar{X}$ and (ii) the sample median is good for μ because it is better than one observation for any symmetric F.

The result in section 3 holds for symmetric F, but is established more generally for any F and its median parameter. Section 4 indicates how the idea of diversification can be reformulated via the median so that, in fact, diversification is always better than putting all eggs in one basket.

2. How Averaging/Diversification can be worse than a Single Observation/Asset

Figure 1 shows densities for a thin (Gaussian) and a fat-tailed situation. While the thin and fat tails are evident, there is nothing in the picture to suggest why the average should perform so differently in the two situations.

Insight comes from looking at bivariate density contours. Figure 2 shows the density contours for the thin-tailed case assuming X_1 and X_2 are independent. Each contour is an iso-probability curve. The figure shows the density contours are circles radiating away from the origin. Also depicted is the half space in the southwest corner showing the (x_1, x_2) where the mean is less than -3. Because the density contours are circles, the most likely values are seen to occur where x_1 and x_2 are approximately equal. When the average takes a value around -3, it is most likely that both x_1 and x_2 are in a neighborhood of -3.

Now consider contours for the fat-tailed case. Many who have not seen such pictures guess that the contours will be circular, and that fat-tailedness will be reflected in the probability values associated with the contours. Density values on the circles far from the origin will be much larger reflecting the greater probability of out-lying tail events.

Figure 3 depicts the density contours for the Cauchy distribution. They are qualitatively very different than the previous picture. The bivariate density is still unimodal with probabilities decreasing in directions away from the origin. But the equi-probable contours are no longer circular, not even convex, becoming ever more concave away from the origin. Instead of probability radiating away from the origin as circular contours, the fat-tailed probabilities accumulate along the axes while becoming sparse in the interior regions of the quadrants. Probabilities are higher along the axes where one of the x_i is large and the other small. In the interior regions, with x_1 and x_2 comparable in magnitude, events are rare.

The concave contours away from the origin are a defining feature of the fat-tailed case. The shape begins to suggest what can go wrong with the average as well as its diversification counterpart.

As in the thin-tailed picture, Figure 3 depicts the half space such that the average is less than -3. The shading shows the (x_1,x_2) that make the average less than -3. These high probability regions are now seen to fall along the axes, away from the interior. The average is most likely less than -3, not when both are around the -3 value, but when one of the observations is less than -6 and the other is less than zero (so that the average is -3). In the fat-tailed case when the average is extreme it is most likely due to only a few or one of the x_i.

Figure 4 shows the sets of (x_1,x_2) such that X^* and $\bar{X}$ are in a neighborhood of the origin. The concave density contours begin to convey how X^* could outperform the average. The probabilities for the average are coming from the sample space where x_1 and x_2 are comparable, and with fat-tailed F these are low probability events. In contrast X^* with all its weight on X_1 is getting its probability mass from along the axis where probability is accumulating.

Figure 5 shows a similar story for a discreet case. In this case X is simply a symmetric version of the St. Petersburg lottery: X takes values $\pm 2^i$ with probability, $\frac{1}{2}2^{-i}$, i=1,2,….. Let X_1,X_2 be independent with common distribution X. The sample space is now a lattice, a portion of which is shown in Figure 5. As in the previous figure the lines connect points of equal probability. The distinctive feature is again the concave equi-probability curves in each quadrant. When $\bar{X}$ is large it is most likely due to one extreme observation.

Besides motivating what can go wrong with the average and its diversification counterpart, the figures suggest the different data patterns generated by thin and fat tailed data. Instead of distinguishing thin and fat tails on the basis of hard to intuit tail integrals we now look for situations in which the average is determined by only a few large magnitude observations. For additional discussion of data patterns underlying fat-tails see Ibragimov (2006) and Neslehova et al. (2006).

3. The Median is better than *X** for any *F*.

For symmetric F there is no ambiguity about the population location: the location estimators, the sample mean, median, and X* all estimate the same population quantity[3]. In this case the sample mean is not good because there exist (symmetric F) such that the mean's sampling distribution is more dispersed than X*'s distribution. The result in this section will show that the situation for the sample median is different: the sample median is always more concentrated than X*.

The result however will be established for general, not necessarily symmetric, F. The sample median is in general an estimator for the population median. The one-observation estimator X* is also an estimator for the median of F since its sampling distribution (namely, F) has a median that is trivially the same as the median of F. The theorem will show that the sample median is always more concentrated about the population median than X*, for any F and hence is inherently good.

The (set of) median(s) of *X* satisfy, $F(m) \le ½ \le F(m+0)$. Let the interval of (population) medians of *X* be denoted, $[\mu^-, \mu^+]$. We will consider arbitrary $\delta > 0$ neighborhoods of $[\mu^-, \mu^+]$ such that the neighborhood does not already contain all of the mass of X, $\Pr[\mu^- - \delta < X < \mu^+ + \delta] < 1$.

Given *X₁,…Xₙ*, the set of sample medians can be defined as the solution set of the absolute error function, $\rho_n(b) = \sum_{i=1}^{n} | x_i - b |$. The solution set is a point when *n* is an odd number, *n=2r+1*, and the median corresponds to the *r+1* order statistic. When *n* is an even number, *n=2r*, the median is the closed interval bounded by the *r* and *r+1* order statistic.

For present purposes it is sufficient to restrict attention to a definition such that the median is always a point.[4] That is, for any *n*, denote the median, m_n, as

---

[3] A location estimator X' is defined as shift and scale equivariant, that is, $X'(X_1 + a, ..., X_n + a) = X'(X_1, ..., X_n) + a$, and $X'(\lambda X_1, ..., \lambda X_n) = \lambda X'(X_1, ..., X_n)$.

[4] This is not essential and with a slight modification the analysis can be done in terms of a set-valued sample median.

(i) the regular median when n is an odd number, and

(ii) the regular median of the first (or randomly selected) *n-1* observations when *n* is an even number.

We now establish,

Theorem. Let $X_1,\ldots,X_n$, $n\geq 3$, be iid with common distribution *F*. The probability that the median is in a $\delta > 0$ neighborhood of $[\mu^-, \mu^+]$ is greater than the probability that *X** is in a $\delta > 0$ neighborhood of $[\mu^-, \mu^+]$ for any $\delta > 0$ and any *F*.

Proof: Since *X** has the same distribution as *X*, the probability that *X** is in a $\delta > 0$ neighborhood is $F(\mu^+ + \delta) - F(\mu^- - \delta + 0)$, which we will write as, $\Delta_1 + \Delta_2$, where $\Delta_1 = F(\mu^+ + \delta) - ½ > 0$ and $\Delta_2 = ½ - F(\mu^- - \delta + 0) > 0$.

Let $G_n(c)$ denote the probability that m_n<c, where $n=2r+1$. This happens if at least $r+1$ of the observations are less than *c*, or

$\Pr[m < c] = G_n(c) = R_n(F(c))$ where, $R_n(u) = \sum_{j=r+1}^{n} \binom{n}{j} u^j (1-u)^{n-j}$. Let $G_n(c+0)$ denote the probability that $m_n \leq c$. This happens if at least $r+1$ of the observations are less than or equal to *c*, or $\Pr[m \leq c] = G_n(c+0) = R_n(F(c+0))$. The probability that the median is in a $\delta > 0$ neighborhood of $[\mu^-, \mu^+]$ is then given by

$$R_n(F(\mu^+ + \delta)) - R_n(F(\mu^- - \delta + 0)) = R_n(½ + \Delta_1) - R_n(½ - \Delta_2).$$

The proof now follows from, $R_n(½ + \Delta_1) > ½ + \Delta_1$, and $R_n(½ - \Delta_2) < ½ - \Delta_2$; see appendix. The appendix also shows that as n increases the probability that the median is in a $\delta > 0$ neighborhood increases. Hence, it always pays to use more observations. It also shows that as *n* increases the probability that the median is in any $\varepsilon>0$ neighborhood goes to one; convergence holds for any *F*. Rates of convergence do depend on *F*, in particular on the smoothness of *F* at the median; for general conditions, see Knight(1998).

Dependence and Non Identical Distributions. We briefly indicate how the result can be extended to the non iid case and the case of weighted averages and weighted medians.

Let $X_1,\ldots,X_n$ where the X_i are arbitrary, possibly dependent and non identically distributed. The benchmark, single observation/asset is still the dartboard that selects one of the X_i at random; $X^* = X_{i^*}$ where i^* =i with probability n^{-1}. The distribution of X^* is denoted by F^* and is now given by the weighted mixture of marginal distributions, $F^*(c) = \sum_{i=1}^{n} w_i F_i(c)$. Let μ (or in case there is an interval of medians, $[\mu^-, \mu^+]$)be the median of F*. X* is an estimator for μ and its distribution is F*. Is there an estimate for μ that is guaranteed to be more concentrated around μ for any F*.

Instead of the regular median, consider the estimate based on a bootstrap sample. That is, let $i_1^*,\ldots,i_n^*$ where i_j^* takes values $1,\ldots,n$ with probability $n^{-1}$and consider the median of the $X_{i_1^*},\ldots,X_{i_n^*}$. These X_i^* are iid with common distribution F^*, the distribution of X^*. The above theorem now applies so that the bootstrap median is an improvement over X^* for any F^*.

For weighted estimates (corresponding to portfolios with different weights on the asset returns), let $\bar{X}$ denote the (weighted) average, $\bar{X} = \sum_{i=1}^{n} w_i X_i$ where the weights w_i, i=1,…,n; $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$.. In this case the benchmark single observation/asset is $X^* = X_{i^*}$ where i^* =i with probability w_i; that is, the dartboard probabilities are determined by the w_i weights. The distribution of X^* is then the weighted mixture of marginal distributions, $F^*(c) = \sum_{i=1}^{n} w_i F_i(c)$. This dartboard estimate is a random draw from the (weighted) empirical distribution, $\hat{F}^*(c) = \sum_{i=1}^{n} w_i I[X_i < (c)]$. The weighted sample mean and median estimates then correspond to the mean and median of $\hat{F}^*$.

4. Median Diversification

The conventional wisdom says: Do not put all your eggs in one basket. But what to do otherwise? If the alternative is it diversify into a basket that pays the (weighted) average then the conventional wisdom is wrong. The conventional wisdom is wrong because putting all your eggs in one basket can be riskier than the one-egg strategy.

To restore the conventional wisdom requires a redefinition of the alternative to putting all your eggs in one basket. To guarantee risk reduction, requires a derivative product that pays not the average, but the median return of a group of assets. Unlike standard diversification based on the average, this will be more concentrated than the dartboard for any F. Diversification is a good thing, but it means that it always pays to put your eggs in the median rather than single asset basket.

Appendix

$R_n(u)$ can be written $R_n(u) = B_n \int_0^u [t(1-t)]^r \, dt$ where $B_n = \frac{(2r+1)!}{r!r!}$ (for example, David (1981, p5.)). This is the cumulative distribution of a beta random variable, β(r+1,r+1). Results for $R_n(u)$ follow from properties of beta random variables. $R_n(u)$ is continuous from [0,1] to [0,1], $R_n(0) = 0$, $R_n(½) = ½$, $R_n(1) = 1$. The median being better than *X** is a consequence of $R_n(u)$ being greater than u for u> ½, and less than u<½. To see this write $R_n(½ + \delta) - ½ = \delta H_n(\delta)$ where $H_n(\delta) > B_n(½(½ - \delta))^r$ [verify that $t(1-t) > ½(½ - \delta)$ for, $½ - \delta < t < ½$, $½ < t < ½ + \delta$. So, $R_n(½ + \delta) - ½ > \delta$ and $R_n(½ - \delta) - ½ < -\delta$.

The median improves with increasing n because $R_{n+2}(½ + \delta) > R_n(½ + \delta)$, $\delta > 0$. To verify write, $R_{n+2}(½ + \delta) - ½ = B_{n+2} \int_{½}^{½+\delta} t(1-t)[t(1-t)]^r \, dt > ½(½ - \delta) \frac{B_{n+2}}{B_n} R_n(½ + \delta)$. The inequality follows from $\frac{B_{n+2}}{B_n} > 4$.

The mean of a beta random variable, β(r+1,r+1) is ½ and the variance is 1/(4r+6) (for example, Feller(1971 p.50). So, as r goes to infinity, β(r+1,r+1) converges in distribution to ½ or, as $n \to \infty$, $R_n(½ + \Delta) \to 1$ while, $R_n(½ - \Delta) \to 0$. Hence the median is in any $\delta > 0$ neighborhood of $[\mu^-, \mu^+]$ with a probability that goes to one as $n \to \infty$ for any *F*.

References

Artzner, Ph., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent Measures of Risk*,* *Mathematical Finance. 9*, 203–228.

David, H.A.(1981). *Order Statistics*. Wiley. Ames.

Embrechts, P., Neslehova, J., Wüthrich, M.V. (2007). Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness, http://www.math.ethz.ch/%7Ebaltes/ftp/papers.html

Fama, Eugene (1965). The Behavior of Stock Market Prices. *Journal of Business* 38, 34-105.

Feller William (1971), *An Introduction to Probability Theory and its Applications*. Volume 2. Second edition. Wiley, New York.

Ibragimov, Rustam and Johan Walden (2007). The Limits of Diversification When Losses may be Large. *Journal of Banking and Finance,* 31, 2551-2569.

Ibragimov, Rustam (2006), On the Robustness of Economic Models to Heavy Tailedness Assumptions. http://www.economics.harvard.edu/faculty/ibragimov/papers.html

Ibragimov, Rustam (2005). Portfolio Diversification and Value At Risk Under Thick-Tailedness, Harvard Institute of Economic Research Discussion Paper Number 2086, August 2005, http://post.economics.harvard.edu/hier/2005papers/2005list.html

Knight, Keith (1998). Limiting Distributions for L_1 Regression Estimators under General Conditions, *The Annals of Statistics,* v 26, n 2, pp 755-770

Mandelbrot, Benoit and Richard L. Hudson (2004). *The (Mis)behavior of Prices: A Fractal View of Risk, Ruin, and Reward.* New York: Basic Books & London.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business* 36, 394-419.

Neslehova, J., Embrechts, P., Chavez-Demoulin, V. (2006): Infinite mean models and the LDA for operational risk, *Journal of Operational Risk,* 1(1), 3-25.

Taleb, Nassim (2007). *The Black Swan,* Random House, New York.

Figure 1
Thin and Fat Tailed Densities

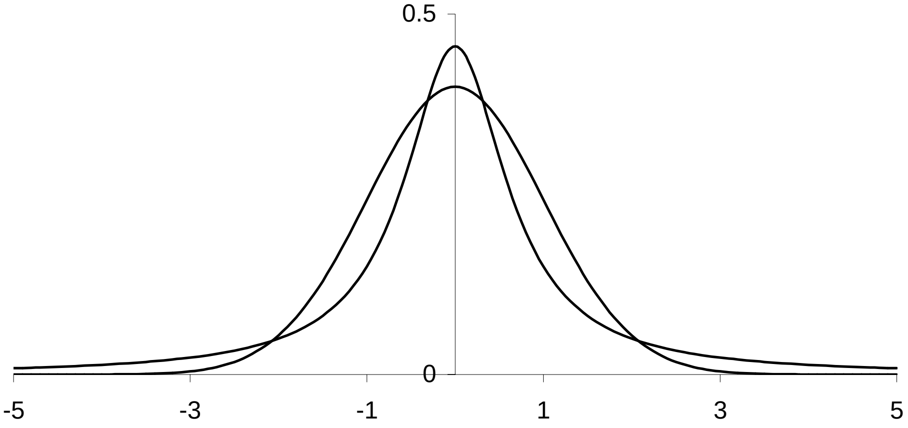

Figure 2
Bivariate Thin-Tailed Density Contours

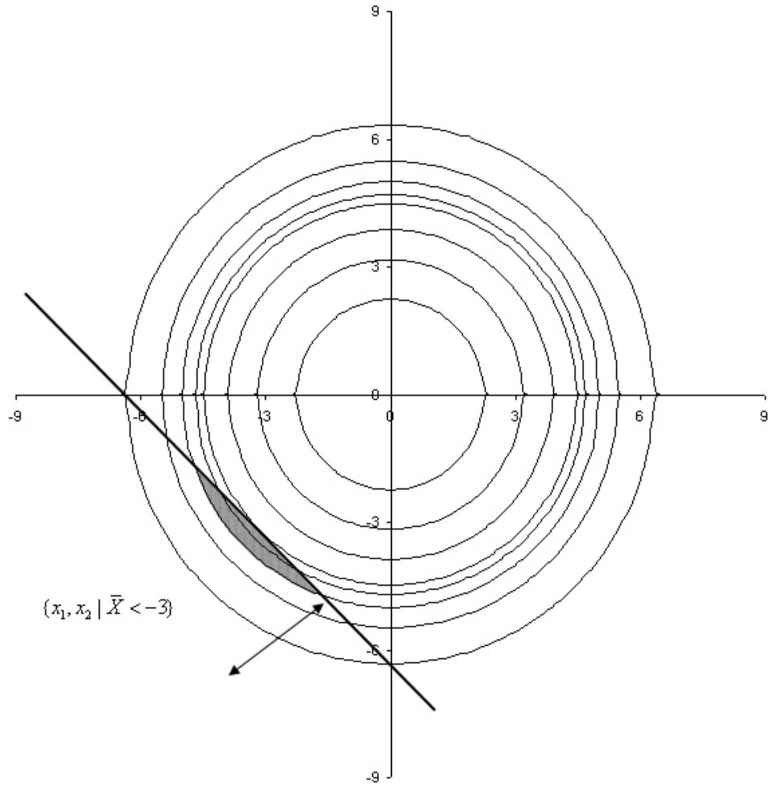

Figure 3
Bivariate Fat-tailed Density Contours

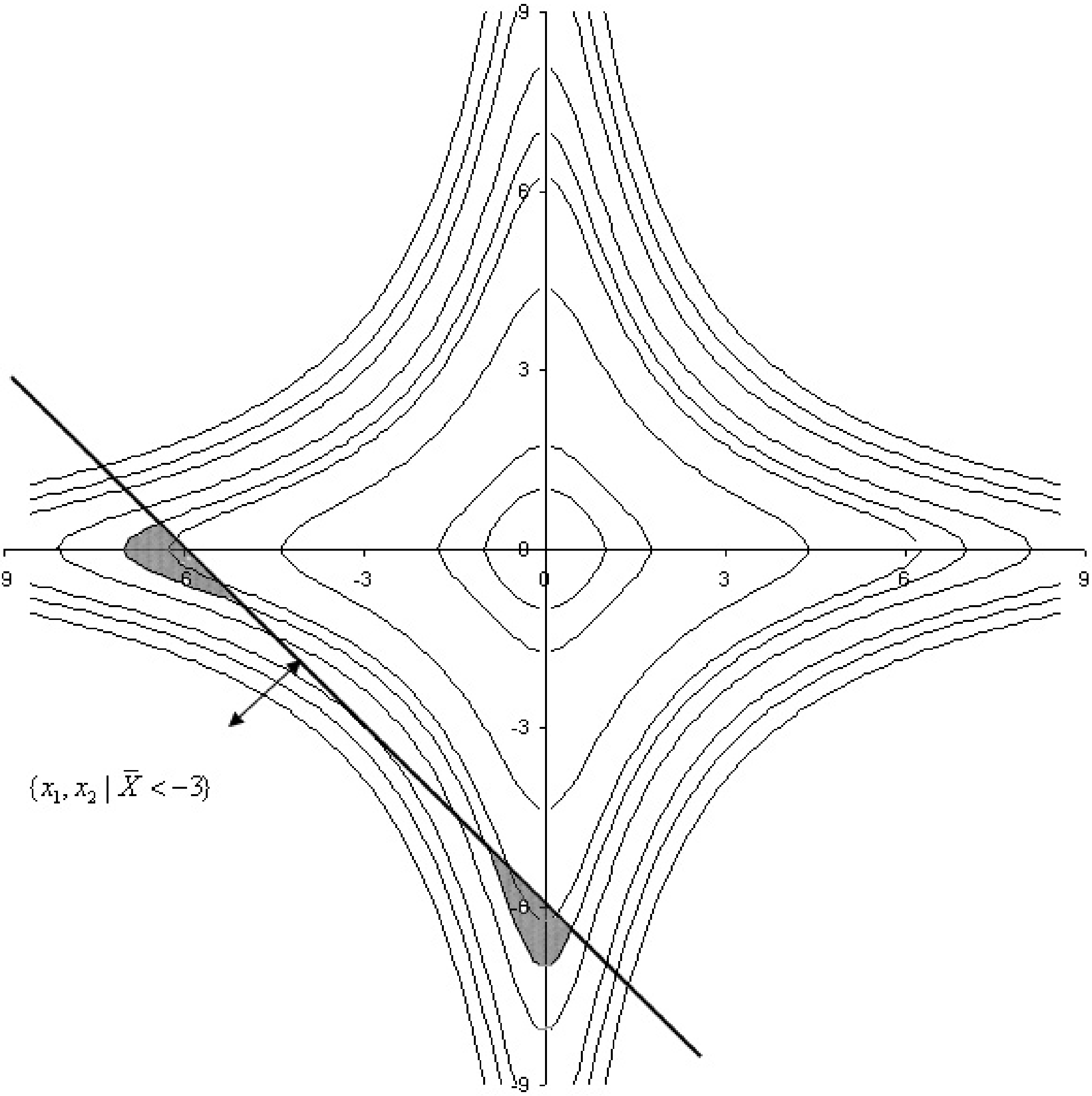

Figure 4

$\delta -$ neighborhoods of μ for X* and $\bar{X}$

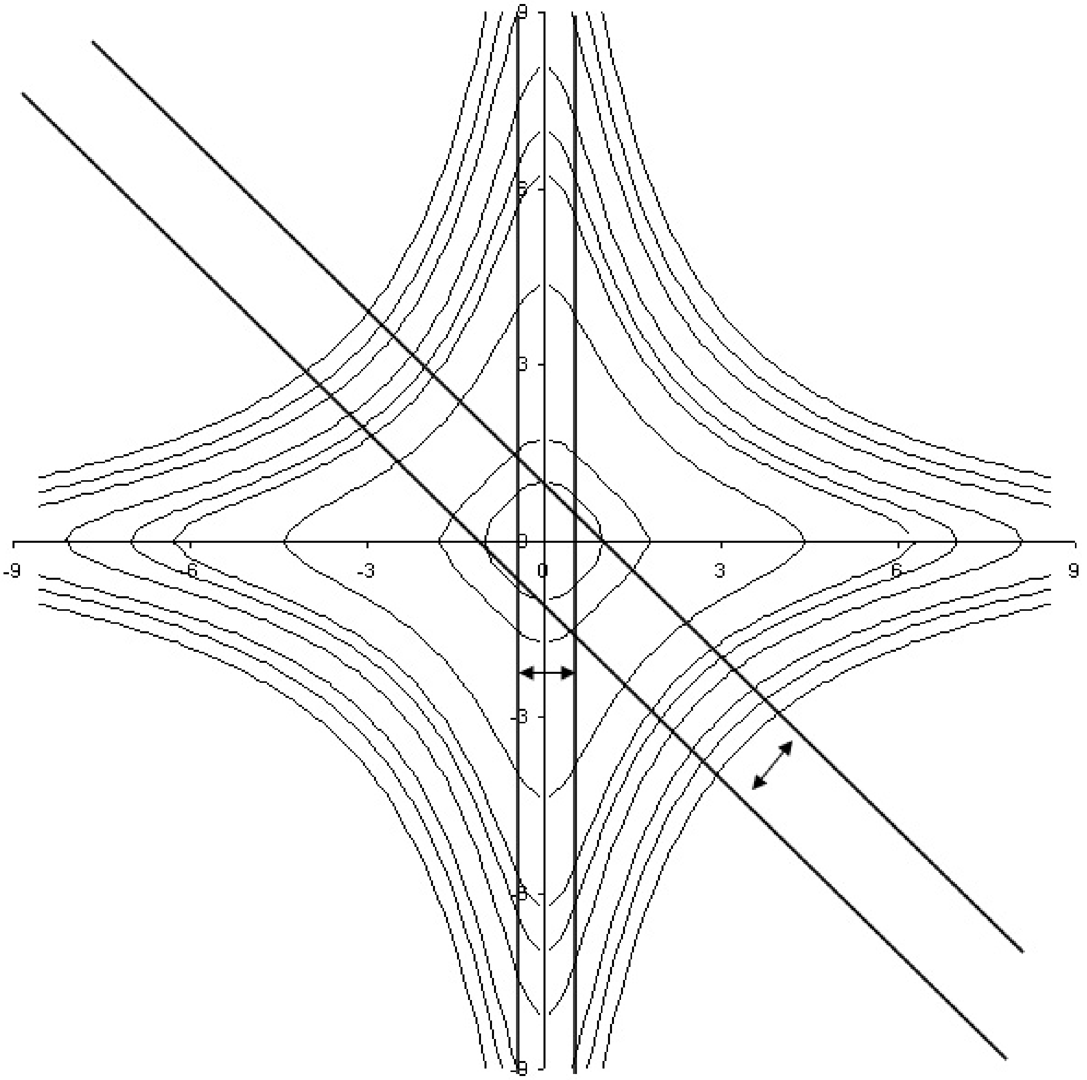

Figure 5
Fat-tailed Discrete Probabilities