Notes on the 'Shape' of Distributions.

Ian Jewitt*
Nuffield College, Oxford

June 2004

Abstract

This paper isolates properties of distributions which are independent of location and scale but which may nevertheless be useful in a variety of economic applications. Particular attention is paid to skewness and other tail orderings. Points of contact are made with the existing statistics literature.

1 Introduction

There has been extensive discussion and use of stochastic dominance concepts to characterise bigger or better random variables and of risk, dispersion etc. to characterise how spread out they are. The aim here is to characterise aspects of the "shape" of a distribution in economically meaningful and analytically tractable ways. We begin by briefly listing a few possible applications.

1.1 Signaling etc.

A classic signaling set up serves to illustrate. Graduating from university is a binary choice and therefore divides the population into two categories. Suppose X measures productivity and the cost of graduating for a type $X = x$ individual $c(x)$ is decreasing in x. In a separating equilibrium where all individuals above some cut-off productivity level

*This paper arose from a number of converstions about different economic issues: I would like to thank Jean Tirole and Roland Benabou who posed the question about the behaviour of the difference $E[X|X \geq t] - E[X|X < t]$ and Meg Meyer and Rebecca Stone who posed the variance question. Meg Meyer drew my attention to the matching context. Jean Tirole bears the chief responsibility of giving me firm encouragement to write things down but he cannot be blamed for the shape of the resulting paper.

t graduate, and where workers are paid their expected probability, the value of the signal is given by $R = E[X|X \geq t] - E[X|X < t]$. The marginal type t must pay this amount in a separating equilibrium. Given a cost $c(t)$ decreasing in t, $E[X|X \geq t] - E[X|X < t] = c(t)$ defines an equilibrium, evidently if R is increasing in t there is only one such equilibrium and it will have regular comparative statics - for instance, reducing the cost of education will increase attendance at university. On the other hand if $E[X|X \geq t] - E[X|X < t]$ is non monotonic, we may have multiple equilibria and perverse comparative statics.

Defining the random variable $S_t = S$ as the indicator of whether $X \geq t$ or not The posterior mean $M = E[X|S]$ is a binary random variable which takes the values $m_0 = E[X|S = 0]$, and $m_1 = E[X|S = 1]$. We are interested therefore in the signaling application of the behaviour of the *range* of the random variable M as a function of t and the shape of the underlying distribution of X.

Benabou and Tirole (2004) also direct attention to the behaviour of the range of M (review 2004 version) in a behavioural model of crowding out of pro social behaviour by financial incentives. Stone (2004) addresses a self-signalling ego utility model in which it is of interest to know whether the variance $var(M)$ is *quasiconvex* or *quasiconcave*.

1.2 Matching

McAfee (2002) considers a situation in which the payoff from a match is increased by the degree of positive association. Specifically, suppose the value of matching $X_1 = x_1$ with $X_2 = x_2$ is given by the *product* $x_1 x_2$. Moreover, assume that X_1 and X_2 are iid. McAfee asks what proportion of the maximum potential expected match value can be secured by a binary classification which partitions the distribution around the mean. That is, one splits both populations into two classes, and then match the good types randomly with the good types and the bad with the bad. Another natural question is how to design the optimal binary classification, or indeed whether one exists. Suppose university serves a purely sorting role for mate selection, should the university system be highly selective or not very? We observe[1] that the value added by the binary classification is simply the variance of M so the question relates to the behaviour of this variance as a function of t and the shape of the underlying distribution.

[1] See the appendix for a proof of this claim.

1.3 The Roy Model

Do markets increase or reduce inequality? Heckman and Honore (1981) discuss this issue (among others) in context of the Roy model of occupational choice in which individuals self select into the occupation in which they are most productive. A somewhat special set up is where individual productivity in different occupations is iid with cdf F_X. Whether market choices reduce inequality below what it would be with random assignment evidently depends on the shape of F_X rather than its location or scale. It can be shown to depend on how fat the tails of the distribution are compared to an identifiable reference distribution.

1.4 Auctions

Imagine Mr Two collected printed ephemera for 50 years in competition with the only other collector Ms One. All sales were conducted by Vickrey auction. Mr Two recorded the price paid for each item he bought but nothing else. Ms One kept no records. Assuming the values for the two bidders are iid across bidders and time, how does Mr Two's records correspond to the underlying distribution of values? In an independent private value Vickrey auction with two bidders, the expected payment by bidder 2 conditional on his winning the auction when he has value $X_2 = x$ is equal to $\varphi(x) = E[X_1|X_1 \leq x] = E_{X_1}[X_1|X_1 \leq x]$. Hence, conditional on bidder 2 winning, his payment is the random variable $\varphi(X_2)$ which has distribution function $F_X(\varphi^{-1}(x)) = \Pr[\varphi(X_2) \leq x] = \Pr[E_{X_1}[X_1|X_1 \leq X_2] \leq x]$ and quantile $\varphi(F_X^{-1}(p)) = E[X_1|X_1 \leq F_X^{-1}(p)] = \frac{\int^{F_X^{-1}(p)} \eta dF_X(\eta)}{p} = \frac{\int^p F_X^{-1}(\zeta)d\zeta}{p}$. Hence, corresponding to a random variable with quantile $F_X^{-1}(p)$ there is another derived random variable with quantile $\frac{\int^p F_X^{-1}(\zeta)d\zeta}{p}$ which has a simple economic interpretation. How are the shapes of the distributions of these two random variables related? In order to compare the valuations of printed ephemera and early 20C oil paintings (with data collected under the same conditions) a researcher displays the QQ plot[2] of one against the other, finds it to be convex and concludes that the data are consistent with her prior belief that the distribution of values for early 20C oils is more skewed to the right than that of printed ephemera. The current paper argues that this is a correct reading of the data. Put differently, we propose criteria for "right skewed" such that if the distribution of valuations of oil paintings is more skewed to the right than for printed ephemera, then for a given bidder so will be the prices paid conditional on winning. A different but

[2] A QQ plot plots for each p the quantiles of the two distributions to be compared. In other words, it is the graph of $F_X^{-1}(F_Y(x))$ where F_X and F_Y are the distribution functions of the two random variables being compared.

related question is whether the sale prices themselves inherit the shape properties of the valuations? For symmetric independent private value Vickrey auctions, this is a classical question about the behaviour of order statistics.

1.5 VAR, Value at Risk, Inequality

VAR represents how bad things get with some small chance. A VAR measure therefore supplements, say mean and standard deviation as characteristics of a distribution. The most common measure is simply the 5% quantile, i.e. $F_X^{-1}(p)$ with $p = 0.05$. This reports the best that can happen conditional on the worst 5% of outcomes occurring. Some authors (refs) have proposed $E[X_1|X_1 \leq F_X^{-1}(p)]$ as a better measure, being the expected outcome conditional on the worst 5% of outcomes occurring.

2 Skewness

2.1 pdf crossings

An intuitive representation of skewness is expressed in terms of the densities. The more right skewed density should have more of a right tail and less of a left tail and this property should not depend on the location and scale of either density. Figure 1 displays a plot of a standard normal density together with a number of different location and scale lognormal densities. The lognormal density with the leftmost mode crosses the normal density three times, so do those with the second and third leftmost nodes, the rightmost lognormal density crosses the normal only once. One way therefore of expressing lognormal distributions more right skewed than normal distributions is by the way these graphs are interlaced as characterised by this pattern of crossings. Specifically, if for any pair of densities f_{aX+b}, $f_{\alpha Y+\beta}$ ($\alpha \neq 0$, $a \neq 0$) selected from the location scale family of distributions associated with the random variables X and Y, f_{aX+b} crosses $f_{\alpha Y+\beta}$ at most 3 times and if three crossings do occur the first is from $f_{aX+b} < f_{\alpha Y+\beta}$ to $f_{aX+b} > f_{\alpha Y+\beta}$, then we assert X to be more right skewed than Y.

2.2 cdf and quantile crossings

There is a more general formulation of the same idea which may be slightly less visually intuitive but which leads to a very useful criterion. This can be derived as a consequence of the above density condition via the variation diminishing property of integration. If any integrable function $g(x)$ crosses the x-axis m times in the interval $[c, d]$ then the integral $\int_c^x g(\eta)d\eta$ will also cross the x-axis at most m times, possibly

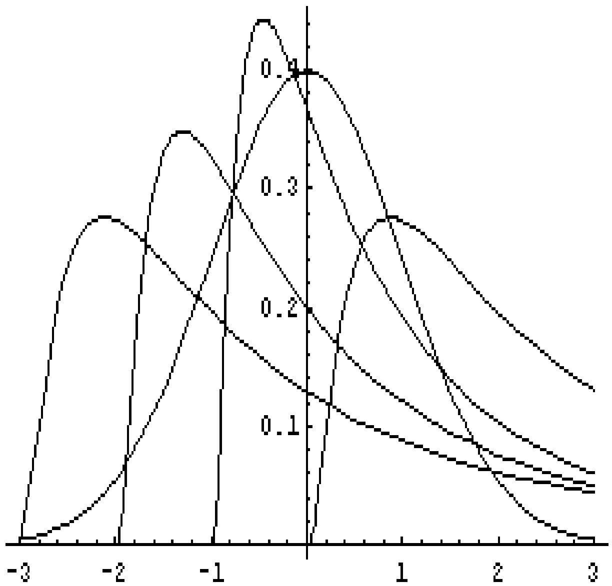

Figure 1: One normal density with four lognormal: at most three crossings.

less. Figure 2 illustrates the graph of a function on $[1, 9]$ with its integral. The graph of the original function crosses the $x$axis seven times, but the integral crosses only six times. Note that if g is the difference of two densities with supports in some interval $[c, d]$ (one or both of c or d can be infinite), $g(x) = f_{aX+b}(x) - f_{\alpha Y+\beta}(x)$ then $g(b) = 0$. This $g(d) = 0$ counts as a sign change for the purpose of this exercise so we are lead to the conclusion that if $f_{aX+b}(x) - f_{\alpha Y+\beta}(x)$ has at most three sign changes, then the difference between the distribution functions can have at most two sign changes, this leads to a very convenient representation. Evidently, iff two functions cross m times, then the inverses (assuming they exist) of the two functions will also cross m times. Hence, setting $a = 1$, $b = 0$ for simplicity (it makes no difference) if the distribution function of $F_X(x)$ crosses $F_Y(\frac{x-\beta}{\alpha})$ at most twice, then $F_X^{-1}(p)$ crosses $\alpha F_Y^{-1}(p) + \beta$ at most twice. Since this is required for each α, β and since the first sign change must be from negative to positive the condition is easily seen to be that $F_X^{-1}(p)$ is a convex transformation of $F_Y^{-1}(p)$. This convexity characterisation of "more right skewed" has already been introduced into the statistics literature by van Zwet (1964). It is worth noting that this relation defines some interesting classes of distributions, for instance distributions which are less right skewed than exponential by this criterion have log concave cdf's.

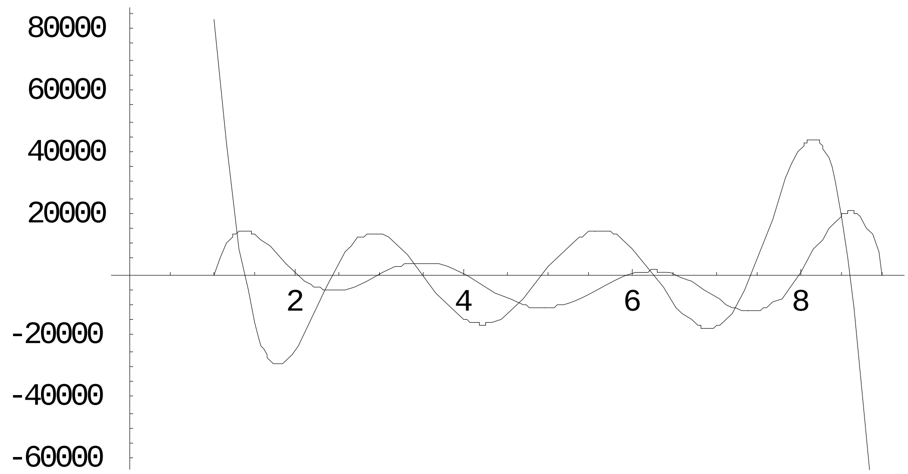

Figure 2: The variation diminishing property: The graph of a function and it integral.

2.3 Integrated cdf and quantile crossings

We can pursue this line of enquiry further and integrate again to derive a yet more general skewness relation. Suppose we choose $\alpha = 1$ without loss of generality and then β, a, b so that both distributions $aX + b$ and $\alpha Y + \beta$ have the same mean, then the area under the quantile functions is equal for both distributions (equal to the zero mean). The strict variation diminishing argument we used above holds again so we can conclude that $\int_0^p (F_X^{-1}(\rho) - \mu_X) d\rho$ and $\beta \int_0^p (F_Y^{-1}(\rho) - \mu_Y) d\rho$ have at most a single sign change, which if it occurs is from negative to positive for each value of $\beta > 0$. In other words, since $\int_0^p (F_Y^{-1}(\rho) - \mu_Y) d\rho < 0$ the ratio

$$\frac{\int_0^p (F_X^{-1}(\rho) - \mu_X) d\rho}{\int_0^p (F_Y^{-1}(\rho) - \mu_Y) d\rho} \tag{1}$$

is nondecreasing. This criterion speaks to the questions posed in section 1.1-1.2. Indeed, we show below that the monotonicity of the ratio (1) is equivalent to the monotonicity of the ratio (2)

$$\frac{E[X|X \geq t] - E[X|X < t]}{E[Y|Y \geq t] - E[Y|Y < t]} = \frac{Range(M_X)}{Range(M_Y)} \tag{2}$$

In this way we show that a consequence of the van Zwet convexity condition and hence also the triple crossing condition on densities is that the ratio (2) is nondecreasing in t. Furthermore, the monotonicity of (1) and (2) is equivalent to the monotonicity of the ratio[3]

$$\frac{Var(M_X)}{Var(M_Y)}. \tag{3}$$

It follows that finding a distribution Y such that $Range(M_Y)$ is independent of t means that those distributions more right skewed than Y will have $Range(M_X)$ increasing. Uniform distributions satisfy this role. Similarly, finding a distribution Y such that $Var(M_Y)$ is constant means that all those distributions X more right skewed than Y will have $Var(M_X)$ increasing. Student t distributions with 2 degrees of freedom satisfy this role.

Consider the auction example in section 1.4. Suppose F_X is right skewed relative to F_Y in the van Zwet $c-$ordering. This means that for any a, b $F_X^{-1}(p) - a - bF_Y^{-1}(p)$ has at most 2 changes of sign, hence by the variation diminishing property so does

$$\int_0^p F_X^{-1}(\rho) - a - bF_Y^{-1}(p)d\rho,$$

and since this function is continuous in p and probabilities do not change sign so does

$$\frac{1}{p}\int_0^p F_X^{-1}(\rho) - a - bF_Y^{-1}(p)d\rho = \frac{\int_0^p F_X^{-1}(\rho)d\rho}{p} - a - b\frac{\int_0^p F_X^{-1}(\rho)d\rho}{p}.$$

Hence, the quantile of $\varphi(X)$ defined in section 1.4 is a convex transform of the equivalent quantile based on Y. This is the result about QQ plots referred to in section 1.4.

2.3.1 Dual Condition

Atkinson (1971) showed that the more risk averse relation[4] of Rothschild and Stiglitz (1968) had a dual counterpart in terms of the Lorenz curves. Specifically (whether for equal mean distributions or not),

$$\begin{gathered}\int_{-\infty}^x F_X ds \ \geq \ \int_{-\infty}^x F_Y ds \text{ for all } x \in (-\infty, \infty) \\ \Longleftrightarrow \\ \int_0^p F_X^{-1} ds \ \leq \ \int_0^p F_Y^{-1} ds \text{ for all } p \in (0, 1).\end{gathered}$$

We shall establish an extension of Atkinson's theorem which is useful in this context.

[3]Note the modified notation, M_X is what we had previously called M, M_Y is of course just the counterpart for Y.

[4]First analysed by Hardy Litlewood and Polya (1929).

Lemma 1 Let F_X and F_Y be distribution functions such that $\int^x F_X ds$ and $\int^x F_Y ds$ have exactly n crossings, then $\int^p F_X^{-1} ds$ and $\int^p F_Y^{-1} ds$ have exactly n crossings. (Note that the sign of the first difference is reversed).

Proof. See Appendix. ■

A corollary is that $\int_0^p F_X^{-1}(\rho) - a - bF_Y^{-1}(\rho)d\rho$ having at most two crossings for each a, b is equivalent to the integrated cdf's $\int_0^z F_X(\eta)d\eta$ and $\int_0^z F_{a+bY}(\eta)d\eta$ having at most two crossings.

2.4 Summary of the hierarchy of relations

To summarise. We have introduced a number of skewness relations, X is more skewed to the right than Y if:

pdf3 Location scale families of densities f_{aX+b} cross $f_{\alpha Y+\beta}$ at most three times. Whenever there are exactly three crossings, $f_{aX+b}(x)$ initially crosses $f_{\alpha Y+\beta}(x)$ from below.

cdf2 Location scale families of distribution functions F_{aX+b} cross $F_{\alpha Y+\beta}$ at most two times. Whenever there are exactly two crossings, $F_{aX+b}(x)$ initially crosses $F_{\alpha Y+\beta}(x)$ from below.

c-order (van Zwet's c-order) The quantile function of X is a convex transformation of the quantile function of Y, i.e. the QQ plot $F_X^{-1}(F_Y(p))$ is convex.

ac-order $\frac{\int_0^p F_X^{-1}(\eta)d\eta}{p}$ is a convex transformation of $\frac{\int_0^p F_Y^{-1}(\eta)d\eta}{p}$

E $\frac{\int_0^p \mu_X - F_X^{-1}(\eta)d\eta}{\int_0^p \mu_Y - F_Y^{-1}(\eta)d\eta}$ is nondecreasing.

As pointed out above *c-order* and *cdf2* are equivalent. There are, of course alternative representations for all these relations: Corresponding to *c-order* is a derivative ratio condition similar to the monotone likelihood ratio test for one distribution function to be a convex transformation of another. Since the derivative of $F_X^{-1}(p)$ can be written $1/f_X(F_X^{-1}(p))$,this can be written as a 'monotone density-quantile ratio' condition reminiscent of the usual monotone likelihood ratio condition[5]

MfQR $\frac{f_X(F_X^{-1}(p))}{f_Y(F_Y^{-1}(p))}$is decreasing

[5]See Parzen (??) for a discussion pof the density quantile function (fQ function) and list of functional forms for common distributions. We observe that concavity of the fQ function is equivalent to logconcavity of the density.

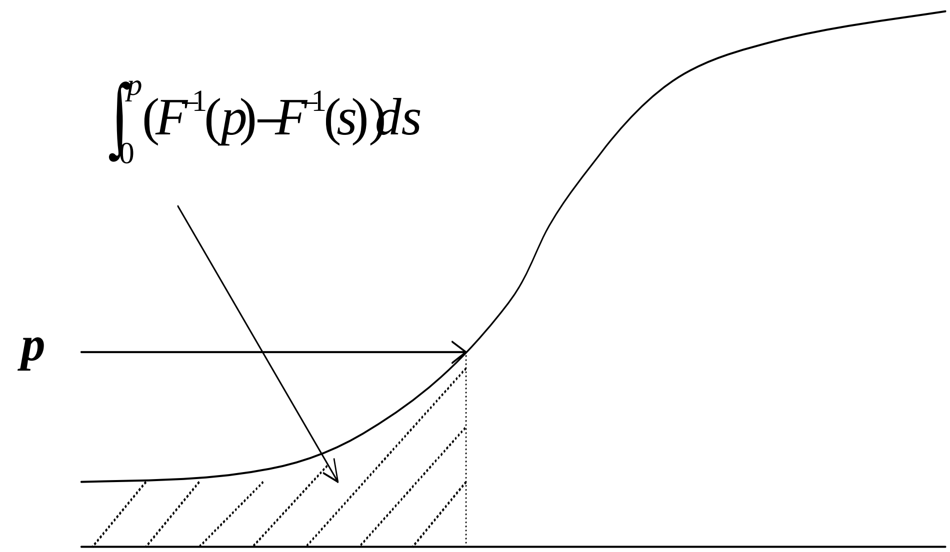

Figure 3:

Similarly, corresponding to *ac-order*, one can derive a monotone derivative ratio test for the convexity order and also the "Arrow-Pratt coefficient of risk aversion" test. These criteria can respectively be written as

$\frac{\int^{F_X^{-1}(p)} F_X(\eta)d\eta}{\int^{F_Y^{-1}(p)} F_Y(\eta)d\eta}$ is increasing.

$\int_0^p \frac{f_X(F_X^{-1}(p))}{f_X(F_X^{-1}(\eta))}\eta d\eta \geq \int_0^p \frac{f_Y(F_Y^{-1}(p))}{f_Y(F_Y^{-1}(\eta))}\eta d\eta$

Note that $\int^{F_X^{-1}(p)} F_X(\eta)d\eta$ is the left tail area of the cdf collected according to percentiles. For a right skewed distribution the condition therefore is that the tail area so measured under distribution X decreases relative to that of Y as we move up the percentiles of the distributions.

Other dual equivalences follow easily from proposition 2.3.1.

Scalar measures of skewness

If, for example, *ac-order* and E really are skewness relations, then if they hold X should be more right skewed than Y by the familiar scalar measures. The skewness of a random variable is often measured by the ratio of the third central moment and cubed standard deviation, or equivalently by the homogeneous version

$$\frac{\sqrt[3]{\int (x-\mu)^3 dF_X(x)}}{\sqrt{\int (x-\mu)^2 dF_X(x)}}.$$

Denote

$$\sigma_{X,m} = \sqrt[m]{\int (x-\mu)^m dF_X(x)}$$

then this can be written

$$\frac{\sigma_{X,3}}{\sigma_{X,2}} \geq \frac{\sigma_{Y,3}}{\sigma_{Y,2}}$$

whenever X is more skewed to the right than Y. One would also expect measures such as

$$\frac{\sigma_{X,2n+1}}{\sigma_{X,2}}$$

for $n = 1, 2, ...,$ to measure skewness. Indeed, for any $k, n = 1, 2, ...,$ with $n \geq k$,

$$\frac{\sigma_{X,2n+1}}{\sigma_{X,2k}}$$

should also intuitively measure skewness.

scalar For all $n \geq k \geq 1$

$$\frac{\sigma_{X,2n+1}}{\sigma_{X,2k}} \geq \frac{\sigma_{Y,2n+1}}{\sigma_{Y,2k}}.$$

We have

Proposition 1 *The following implications hold*

$$\textit{pdf3} \implies \textit{c-order} \implies \textit{ac-order} \implies E \implies \textit{scalar}$$

Proof. The proof of all but the last implication is in the text. See the appendix for a proof of $E \implies \textit{scalar}$. ■

3 Moments of Conditional Means.

Given the family of random variables

$$S_t = \begin{matrix} 1 \text{ if } X \geq t \\ 0 \text{ if } X < t. \end{matrix}$$

and

$$M_t = E[X|S_t].$$

Evidently, by the law of iterated expectations M_t has mean $E[M_t] = E[X] = \mu_X$. Other moments of the distribution are characterised in the following lemma. .

Lemma 2 Let X have distribution function F_X and quantile function F_X^{-1}.

1. M_t has n'th central moment σ_t^n given by

$$\begin{aligned}(\sigma_t^n)^{\frac{1}{n}} &= (E\left[(M_t - E[M_t])^n\right])^{\frac{1}{n}} \\ &= V_n(p)\int_0^p \mu_X - F_X^{-1}(s)ds\end{aligned}$$

2. M_t has range $M_1 - M_0 =$

$$E[X|S_t = 1] - E[X|S_t = 0] = ((1-p)p)^{-1}\int_0^p \mu_X - F_X^{-1}(s)ds$$

where $p = F_X(t)$ and

$$V_n(p) = \left(p^{1-n} + (-1)^n(1-p)^{1-n}\right)^{\frac{1}{n}}.$$

Proof. Straightforward. See appendix. ■

Throughout, we are only interested in ordinal properties of these moments, specifically monotonicity, quasiconvexity and quasiconcavity. It will suffice therefore to study the ordinal properties of the ratios

$$\frac{\int_0^p \mu_X - F_X^{-1}(s)ds}{\int_0^p \mu_Y - F_Y^{-1}(s)ds}$$

and for *even* moments to look for distributions Y for which

$$\left(p^{1-n} + (1-p)^{1-n}\right)^{\frac{1}{n}}\int_0^p \mu_Y - F_Y^{-1}(s)ds$$

is constant. Note that it is trivial to characterise such distributions in terms of their quantiles $F_Y^{-1}(p)$ and in terms of their "fQ" functions[6] $f_Y(F_Y^{-1}(p))$ by differentiation. Figure 3 shows the fQ functions which achieve this constancy (generated by Mathematica) for $n = 2$, 2.5, 3 and 4. $n = 2$ corresponds to the unimodal density with the highest value at $p = 0.5$, lower values at $p = 0.5$ correspond to higher values of n.

3.1 Distributions having constant range and variance of M_t

Let F_Y be uniform on $[0, 1]$, then

$$\begin{aligned}\int_0^p \mu_X - F_Y^{-1}(s)ds &= \int_0^p (0.5 - s)\,ds \\ &= \frac{p(1-p)}{2}.\end{aligned}$$

[6] See Parzen (??) for a discussion and list of functional forms for common distributions. We observe that concavity of the fQ function is equivalent to logconcavity of the density.

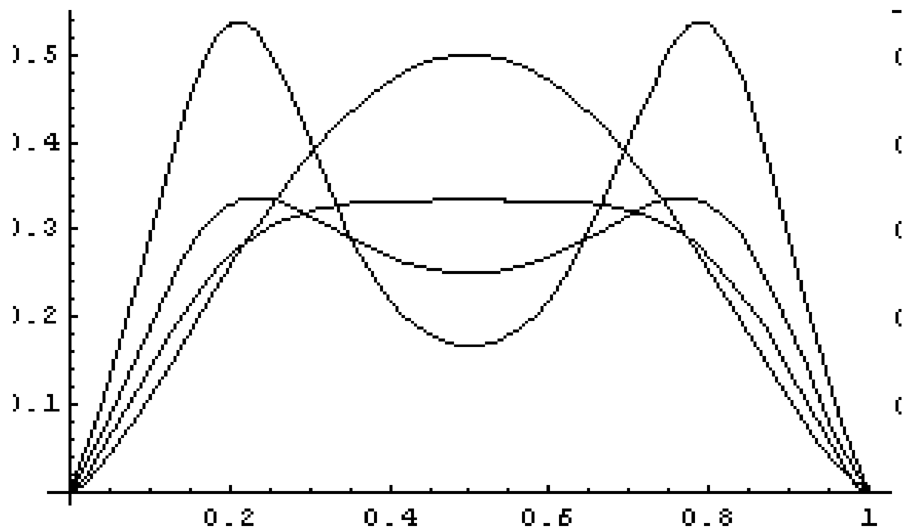

Figure 4: fQ plots for $n = 2, 2.5, 3.4$.

Hence, any distribution which is more right skewed than uniform distributions by any of the criteria introduced above will exhibit $E[X|X \geq t] - E[X| < t]$ nondecreasing in t.

Now we look for a distribution F_Y such that

$$\begin{aligned}\int_0^p \mu_X - F_Y^{-1}(s)ds &= k/V_2(p) \\ &= k/\sqrt{p(1-p)}\end{aligned}$$

for some constant k. Differentiating yields

$$F_Y^{-1}(p) = \mu_X + \frac{1}{2}\frac{1-2p}{\sqrt{p(1-p)}}$$

Equating this to x, we obtain a quadratic equation in p which can be solved to give

$$p = F_Y(x) = \frac{1}{2}(1 + \frac{x}{\sqrt{1+x^2}}),$$

the density is

$$f_Y(x) = 0.5\left(1+x^2\right)^{-\frac{3}{2}}.$$

This is a t-distribution with 2 degrees of freedom. Hence, a t-distribution with 2 degrees of freedom has $var(M_t)$ independent of the partition t. Hence, any distribution more right skewed than such a t-distribution will exibit an increasing variance of the conditional mean M_t as t increases.

3.2 More on Skewness.

This condition is convenient for a number of purposes, for instance, by Jensen's inequality the average of an n-sample drawn from distribution X has the same distribution as an n-sample of convex transformations of Y

$$\frac{1}{n}\sum X_i \stackrel{d}{=} \frac{1}{n}\sum \varphi(Y_i) \geq \varphi(\frac{1}{n}\sum Y_i)$$

hence $\frac{1}{n}\sum X_i$ stochastically dominates $\varphi(\frac{1}{n}\sum Y_i)$. Therefore, denoting F_n as the distribution function of the n sample average of F

$$F_{Xn}(x) \leq F_{Yn}(\varphi^{-1}(x))$$

i.e. the QQ plot of the sample averages lies above the QQ plot of the underlying distributions. Using Jensen's inequality again,

$$\begin{aligned}\int F_X^{-1}(F_Y(y))dF_Y(y) &\geq F_X^{-1}(F_Y(\int ydF_Y(y))) \\ &= F_X^{-1}(F_Y(\mu_Y))\end{aligned}$$

but since

$$\begin{aligned}\int F_X^{-1}(F_Y(y))dF_Y(y) &= \int_0^1 F_X^{-1}(p)dp \\ &= \int xdF_X(x) = \mu_X\end{aligned}$$

we have

$$F_X^{-1}(F_Y(\mu_Y)) \geq \mu_X$$

or

$$F_Y(\mu_Y) \geq F_X(\mu_X).$$

Suppose therefore that X is symmetric about its mean, then Y has more than half its probability mass above its mean.

If the skewness property is of interest, it will be of interest to know under what statistical operations it is preserved. We have already argued that if X_1 and X_2 are iid and right skewed relative to the *iid* pair Y_1 and Y_2 then $E_{X_1}[X_1|X_1 \geq X_2]$ is right skewed relative to $E_{Y_1}[Y_1|Y_1 \geq Y_2]$. It is known that

Proposition 2 (Van Zwet) *Order statistics preserve the van Zwet c-order skewness relation.*

Proof. Immediate from the fact that the k'th order statistic from an $n-$sample of distribution F is given by

$$\begin{aligned} F_{k:n}(z) &= H_{k:n}(F(z)) \\ &= \sum_{j=k\ldots,n} C(n,j)\,[F(z)]^j\,[1-F(z)]^{n-j} \end{aligned}$$

Hence, $F_{k:n}^{-1}(G_{k:n}^{-1}(p)) = F^{-1}(G^{-1}(p))$. ■

Hence, in an independent (iid) private value Vickrey auction the sale price will be right skewed if the valuations are right skewed.

The following is a natural result. Recall, a logconcave (PF_2) density f satisfies

$$\begin{vmatrix} f(x_1-y_1) & f(x_1-y_2) \\ f(x_2-y_1) & f(x_2-y_2) \end{vmatrix} \geq 0$$

for all $x_1 < x_2 < x_3$, $y_1 < y_2 < y_3$, log concave densities are also known as Polya Frequency functions of order 2 (PF_2) a PF_3 density satisfies this condition together with

$$\begin{vmatrix} f(x_1-y_1) & f(x_1-y_2) & f(x_1-y_3) \\ f(x_2-y_1) & f(x_2-y_2) & f(x_2-y_3) \\ f(x_3-y_1) & f(x_3-y_2) & f(x_3-y_3) \end{vmatrix} \geq 0.$$

Proposition 3 *(in the literature?, PF_3 needed?) If X_2 and Y_2 both have PF_3 densities and X_1 and X_2 are right skewed relative to Y_1 and Y_2 respectively (either according to c-order or ac-order) then $X_1 + X_2$ is right skewed relative to $Y_1 + Y_2$.*

(Outline Proof: $F_X(x) - F_Y(\frac{x-a}{b})$ has ≤ 2 sign changes for each a, b. Therefore given the PF_3 assumption so does $\int \left(F_{X_1}(x) - F_{Y_1}(\frac{x-a}{b})\right) F_{X_2}(x-dy) = F_{X_1} * F_{X_2} - F_{Y_1} * F_{X_2}$, where the $*$ denotes the convolution operator. Hence, inverting, $(F_{X_1} * F_{X_2})^{-1}$ is more convex than $(F_{Y_1} * F_{X_2})^{-1}$. Similarly, $\int \left(F_{Y_2}(\frac{x-a}{b}) - F_{X_2}(x)\right) F_{Y_1}(x-dy) = F_{Y_2} * F_{Y_1} - F_{X_2} * F_{Y_1}$ has at most two sign changes (in the opposite order) so $(F_{X_2} * F_{Y_1})^{-1}$ is more convex than $(F_{Y_2} * F_{Y_1})^{-1}$ the result now follows from the transitivity of the 'more convex' relation and the symmetry of the convolution operator.)

Similarly, we have

Proposition 4 *If X_2 and Y_2 both have logconcave densities and X_1 and X_2 are right skewed relative to Y_1 and Y_2 respectively (according to condition E) then $X_1 + X_2$ is right skewed relative to $Y_1 + Y_2$ (according to condition E).*

Proof. Similar to the above but we only need preserve the single sign change. ■

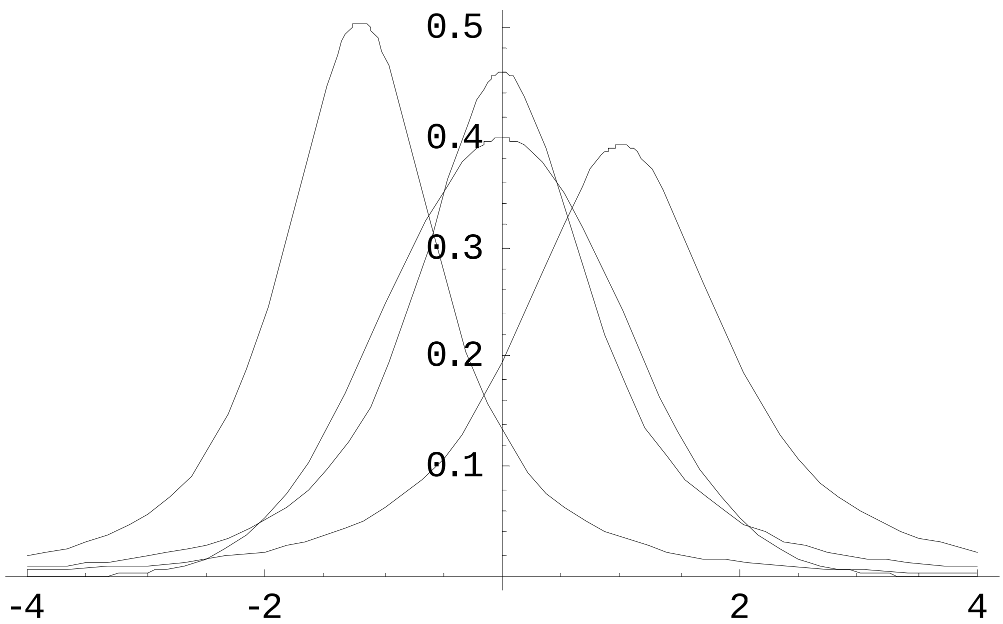

Figure 5: One normal density with three t-distributions: at most four crossings.

4 Fatter tails

The same methods apply. Rather than, say, require a fatter right tail and a thinner left tail for right skewness, we require both tails to be fatter - in a sense. In terms of the density comparison, we require each pair of densities from the respective location scale equivalence classes to have at most four sign changes. Figure 3 displays four t-distributions each with 2 degrees of freedom but with different locations and scales, there is also a normal distribution. One can easily spot the normal, it has thinner tails. One way to quantify "thinner tails" in this figure is to note that the graph of the normal never crosses the graph of a t-density more than four times. In the figure there are two densities which cross the normal twice and one which crosses four times, the one which crosses four times crosses first from below.

As before, one can obtain a hierarchy of more general relations by integrating and appealing to the variation diminishing property of the integral operator.

pdf4 Location scale families of densities f_{aX+b} cross $f_{\alpha Y+\beta}$ at most four times. Whenever there are exactly three crossings, $f_{aX+b}(x)$ initially crosses $f_{\alpha Y+\beta}(x)$ from above.

cdf2 Location scale families of distribution functions F_{aX+b} cross $F_{\alpha Y+\beta}$

at most three times. Whenever there are exactly two crossings, $F_{aX+b}(x)$ initially crosses $F_{\alpha Y+\beta}(x)$ from above.

s-order The quantile of X is an s-shaped (convex-concave) transformation of the quantile of Y, i.e. the QQ plot $F_X^{-1}(F_Y(p))$ is convex-concave.

as-order $\frac{\int_0^p F_X^{-1}(\eta)d\eta}{p}$ is an s-shaped transformation of $\frac{\int_0^p F_Y^{-1}(\eta)d\eta}{p}$

E $\frac{\int_0^p \mu_X - F_X^{-1}(\eta)d\eta}{\int_0^p \mu_Y - F_Y^{-1}(\eta)d\eta}$ is quasiconvex.

Proposition 5 *The following implications hold*

$$pdf4 \implies cdf3 \implies s\text{-}order \implies as\text{-}order \implies E$$

4.0.1 Van Zwet's S-order

For distributions symmetric about some point c, Van Zwet defined a relation as follows: if $F_X^{-1}(F_Y(x))$ is convex on $(-\infty, c)$ and concave on (c, ∞) then F_X is . The definition can evidently be modified to dispense with the assumption of symmetry which is what we do above. An equivalent expression is that the density-quantile ratio

$$\frac{f_Y(F_Y^{-1}(p))}{f_X(F_X^{-1}(p))} \tag{4}$$

is quasiconcave.

4.0.2 Lawrence's Order

Lawrence defined a generalisation of van Zwet's S-order (for distributions symmetric about their means) by requiring that

$$\frac{F_X^{-1}(F_X(x)) - \mu_X}{x - \mu_Y}$$

be decreasing for $x < \mu_Y$ and increasing for $x > \mu_Y$. In other words, additional to the symmetry,

$$F_X^{-1}(p) - \mu_X - k(F_Y^{-1}(p) - \mu_Y)$$

has at most three sign changes which if three occur start with negative to positive. Hence, by the variation diminishing property

$$\int_0^p (F_X^{-1}(\eta) - \mu_X)d\eta - k\int_0^p (F_Y^{-1}(\eta) - \mu_Y)d\eta$$

has at most two sign changes (first negative)

$$\frac{\int_0^p (F_X^{-1}(\eta) - \mu_X) d\eta}{\int_0^p (F_Y^{-1}(\eta) - \mu_Y) d\eta}$$

is quasiconcave.

Note that by these criteria, any unimodal distribution has heavier tails than a uniform distribution. Hence, for any unimodal distribution, $E[X|X \geq t] - EX|X < t]$ is quasiconvex. For any distribution with heavier (repectively lighter) tails than a t-distribution with 2 degrees of freedom will have $var(M_t)$ quasiconvex (respectively quasiconcave).

5 References

Atkinson (1971), ".....''*Journal of Economic Theory*,

Benabou, R. and J. Tirole (2003) "Incentives and Prosocial Behaviour", Toulouse.

Chateauneuf, A. M. Cohen, , and I. Meilijson (2003) "Four Notions of Mean-Preserving Increase in Risk, Risk Attitudes and Applications to the Rank-Dependent Expected Utility Model", *Journal of Mathematical Economics*,

Hardy, Littlewood and Polya (1929)

Jewitt, I. (1989) "Choosing Between Risky Prospects: the Characterisation of Comparative static Results and Location Independent Risk, *Management Science*, 35, pp. 60–70.

Lawrence, (1966) "Inequalities of s-ordered distributions", *The Annals of Statistics*, 3, 413-428.

McAfee, R. (2002), *Econometrica*,

MacGillivray, H.L. (1986) "Skewness and Asymmetry: Measures and Orderings", *The Annals of Statistics*, vol 14, No. 3, 994-1011.

Oja, H. (1981) "On Location, Scale Skewness and Kurtosis of Univariate Distributions", *Scandinavian Journal of Statistics*, 8, 154-168.

Parzen, (??) "

Rothschild and Stiglitz (1968), "...",

Stone, R (2004) "...", DPhil thesis, University of Oxford.

van Zwet, W.R. (1964) "Convex transformations of Random Variables", Math Centrum, Amsterdam.

6 Appendix

Matching

The expected match value when "successes" $(X \geq t)$ are paired randomly with other successes and "failures" $(X < t)$ with other failures exceeds the expected match value when pairs are matched completely randomly by an amount equal to $var(M)$.

Proof. $var(M) = var(E[X|S]) = E[E[X|S]^2] - E[E[X|S_t]]^2 = E[E[X|S_t]^2] - \mu^2$. Assuming X, X_1 and X_2 are iid $\mu^2 = E[X_1 X_2]$ is the value of unsorted matching and $E[E[X|S_t]^2] = E[E[X_1|X_1 \geq t]E[X_2|X_2 \geq t]]$ is the value of sorted matching. ■

Proof. Proof of Lemma 1.

1 follows immediately from the law of iterated expectations.

The n'th central moment is

$$\sigma_t^n = F_X(t)\left(\frac{\int^t (x-\mu_X)dF_X}{F_X(t)}\right)^n + (1-F_X(t))\left(\frac{\int_t (x-\mu_X)dF_X}{1-F_X(t)}\right)^n.$$

But since $\int (x-\mu_X)dF_X \equiv 0$, $\int_t (x-\mu_X)dF_X = -\int^t (x-\mu_X)dF_X$

$$\begin{aligned}\sigma_t^n &= F_X(t)\left(\frac{\int^t (x-\mu_X)dF_X}{F_X(t)}\right)^n + (1-F_X(t))\left(\frac{-\int^t (x-\mu_X)dF_X}{1-F_X(t)}\right)^n \\ &= \left(\int^t (x-\mu_X)dF_X\right)^n \left(F_X(t)\left(\frac{1}{F_X(t)}\right)^n + (1-F_X(t))\left(\frac{-1}{1-F_X(t)}\right)^n\right).\end{aligned}$$

Setting $p = F_X(t)$ and making a change of variables in the integration gives the result after straightforward rearrangement. The calculation for the range is similar. ■

Skewness measures

Note that x^2 has a constant second derivative, x^3 has an increasing second derivative and the ratio of second derivatives of x^3 and x^2 is increasing. Similarly, x^{2n+1} and x^{2k} has ratio of second derivatives which is increasing whenever $n \geq k$.

Proposition: Suppose that

$$\frac{\int_0^p \mu_X - F_X^{-1}(s)ds}{\int_0^p \mu_Y - F_Y^{-1}(s)ds} \text{ is increasing on } (0,1)$$

then whenever $u(x)$ and $v(x)$ have

$$\frac{v''(x)}{u''(x)} \text{ increasing}$$

any c such that $\int u(c(x-\mu_X))dF_X(x) \geq \int u((y-\mu_Y))dF_Y(y)$ implies $\int v(c(x-\mu_X))dF_X(x) \geq \int v((y-\mu_Y))dF_Y(y)$. Hence, choosing $u(x) = x^{2k}$, $v(x) = x^{2n+1}$ yields the following moment inequalities

$$\frac{\sigma_{X,2n+1}}{\sigma_{X,2k}} \geq \frac{\sigma_{Y,2n+1}}{\sigma_{Y,2k}}.$$

Proof. Integrate by parts twice to obtain

$$\begin{aligned}
&\int u(c(x-\mu_X))dF_X(x) - \int u(x-\mu_Y)dF_Y(x) \\
&= \int u(x)dF_X(\frac{z}{c}+\mu_X) - \int u(z)dF_Y(z+\mu_Y) \\
&= \int u(x)d\left(F_X(\frac{z}{c}+\mu_X) - F_Y(z+\mu_Y)\right) \\
&= \int u''(x)\left[\int^x \left(F_X(\frac{z}{c}+\mu_X) - F_Y(z+\mu_Y)\right)dz\right]dx
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\int v(c(x-\mu_X))dF_X(x) - \int v(x-\mu_Y)dF_Y(x) \\
&= \int v''(x)\left[\int^x \left(F_X(\frac{z}{c}+\mu_X) - F_Y(z+\mu_Y)\right)dz\right]dx
\end{aligned}$$

Now choose constant $a \geq 0$ such that the integrand in the following expression is nonnegative (by choosing a so that the two terms $(v''(x) - au''(x))$ and $\int^x \left(F_X(\frac{z}{c}+\mu_X) - F_Y(z+\mu_Y)\right)dz$ cross at the same point)

$$= \int (v''(x) - au''(x))\left[\int^x \left(F_X(\frac{z}{c}+\mu_X) - F_Y(z+\mu_Y)\right)dz\right]dx.$$

This proves the first part of the proposition. For the moment inequality, choosing c such that

$$c.\sqrt[2k]{\int (y-\mu)^{2k}dF_Y(y)} = \sqrt[2k]{\int (x-\mu)^{2k}dF_X(x)}$$

implies

$$c.\sqrt[2n+1]{\int (x-\mu_X)^{2n+1}dF_X(x)} \geq \sqrt[2n+1]{\int (y-\mu_Y)^{2n+1}dF_Y(y)}.$$

Substituting for c yields the result. ■

Dual Condition

Proof. We utilise the fact that $\Phi = \int^x F_X ds$ and $\Psi = \int^x F_Y ds$ are convex functions and are therefore characterised by their convex conjugates

$$\Phi^*(p) = \max px - \int^x F_X ds, \;\; \Psi^*(p) = \max px - \int^x F_Y ds.$$

Suppose that there are sign changes at $(x_1, x_2, ..., x_n)$ with the first being from negative to positive. For simplicity assume that $\Phi - \Psi$ is differentiable, then in each interval (x_i, x_{i+1}), $\Phi - \Psi$ has a stationary point with $\varphi(x_i') - \psi(x_i') = 0$. By duality, there is some $p_i' = \varphi(x_i') = \psi(x_i')$ such that

$$\Phi^*(p_i') = p_i' x_i' - \int^{x_i'} F_X ds, \;\; \Psi^*(p_i') = \max p_i' x_i' - \int^{x_i'} F_Y ds.$$

and hence $\Phi(x_i') - \Psi(x_i') = -(\Phi(p_i') - \Psi(p_i'))$. Evidently, we have we have $p_1' < p_2' ... < p_n'$ otherwise there would be no point of increase in $\varphi(x_i') = \psi(x_i')$ and no change in sign of $\Phi - \Psi$. Now take the interval $(-\infty, x_1)$, either $\Phi - \Psi$ has a stationary point, or it is increasing throughout the interval with $\varphi > \psi$. Choose $p_0' = \psi(x_0')$ for some x_0' in the interval with $\Phi(x_0') > \Psi(x_0')$, then since x_0' is maximising for $\Psi^*(p_0')$ but not necessarily for $\Phi^*(p_0')$

$$\Phi^*(p_0') \geq p_0' x_0' - \int^{x_0'} F_X ds, \;\; \Psi^*(p_0') = p_0' x_0' - \int^{x_0'} F_Y ds,$$

that is $\Phi^*(p_0') > \Psi^*(p_0')$. A similar construction shows that there is also a sign change in $\Phi^* - \Psi^*$ at some $p_{n+1.}$ Hence, we have shown that if $\Phi - \Psi$ has n sign changes then $\Phi^* - \Psi^*$ has *at least* n sign changes. Since by conjugate duality,

$$\Phi(x) = \max px - \Phi^*(p), \;\; \Psi(x) = \max px - \Psi^*(p),$$

the identical argument establishes that if $\Phi^* - \Psi^*$ has m sign changes then $\Phi - \Psi$ has at least m sign changes. Therefore, $m = n$.

$$\begin{aligned} \Phi^*(p) &= \max_x px - \int^x F_X ds \\ &= p\varphi^{-1}(p) - \int_{-\infty}^{\varphi^{-1}(p)} F_X ds \end{aligned}$$

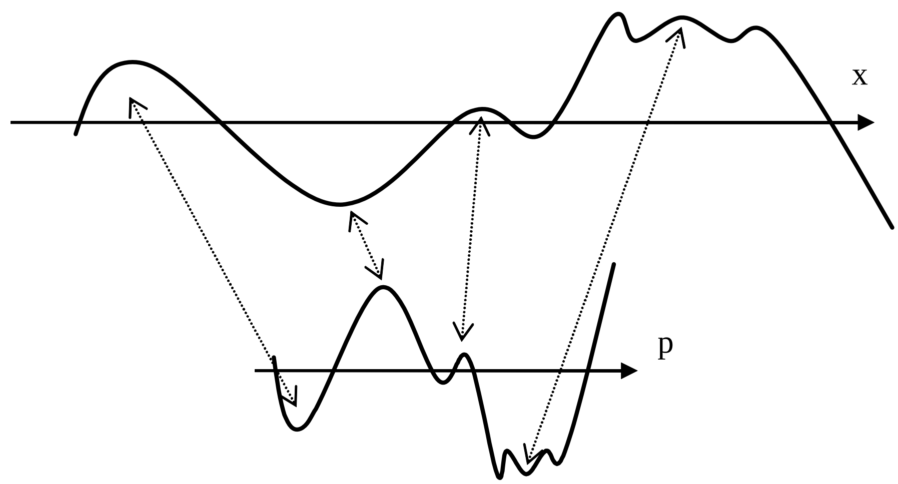

Figure 6: Dual crossing properties of convex conjugates.

from substitution of the first-order conditions. Upon integration by parts, and a change of variable

$$= \int_{-\infty}^{\varphi^{-1}(p)} s dF_X(s)$$
$$= \int_0^p F_X^{-1}(s) ds.$$

■