Psychological Bulletin
1990, Vol. 108, No. 3, 551-567

0033-2909/90/$00.75

Truth and Consequences of Ordinal Differences in Statistical Distributions: Toward a Theory of Hierarchical Inference

James T. Townsend
Indiana University

A theory is presented that establishes a dominance hierarchy of potential distinctions (order relations) between two distributions. It is proposed that it is worthwhile for researchers to ascertain the strongest possible distinction, because all weaker distinctions are logically implied. Implications of the theory for hypothesis testing, theory construction, and scales of measurement are considered. Open problems for future research are outlined.

There are many occasions in psychological research when experimenters are concerned with two or more treatments and their effects on experimental groups. In most instances, the measurements assessing the treatments are assumed to be on at least an ordinal scale. The measurements typically form a sample probability distribution. The investigator may then ask the question as to whether the two distributions differ from one another in some way. The way chosen is often through the testing of summary statistics, particularly sample means, but there are techniques that test entire distributions against one another.

In this article I develop a theory that shows that some differences between the distributions are stronger than others, in that the stronger ones imply the weaker ones but not vice versa. In fact, a dominance hierarchy of distinctions between two arbitrary distributions is developed: A distinction may imply another, be implied by another, be equivalent to another, or none of these. The dominance hierarchy is put into the form of an implication graph. The power of the results is that they are, except for special cases to be specified later, general across distributions. That is, most of the relations do not depend on a particular type of distribution such as the normal distribution. Furthermore, as long as the measurements are on at least an ordinal scale, again with the exception of the special cases, they are scale, and transformation-free up to a strictly monotonic transformation. Thus, in general the distinctions and implications do not depend on interval or ratio level measurement, and investigators are free to perform useful monotonic transformations on their data.

One immediate implication of the theory is that other things being equal, one can seek the strongest distinction reasonable in a given experimental context. Not only does this entail the truth of the weaker distinctions, it also means that the experimental effect holds in a stronger sense. For instance, perhaps the strongest distinction would be that every member (score, measurement, etc.) of the distribution under one treatment would be greater than every member of the distribution under the alternate treatment. This is an extremely strong distinction but is, of course, rarely found in real studies. As I show, a more likely possibility is that for each measurement or score, if the cumulative frequency (distribution function) of, say, Group A is always higher than that of Group B at this score, then these several other weaker distinctions, such as an ordering of the means, are implied. In turn, this distributional ordering is implied by other even stronger distinctions.

Consider the situation in which an experimenter has administered a psychological test to two groups, one presumably pathological and the other presumably "normal." It is possible that the frequency distributions of the test might differ in a very strong way, or they might differ in only a weak way. As a second example, consider a situation in which the experimenter is examining performance through reaction time or accuracy, when the subjects are under two different levels of information processing load. The effect of the increased load on the underlying processing mechanisms could be quite strong or, conversely, might have a much weaker effect. Such consequences might be of importance not only in terms of the practical implications for real-time performance, but also for their implications with respect to potential theoretical explanations such as quantitative process models. Thus, the ramifications for the type and degree of capacity limitation may hinge on the level at which the experimental distinction exists.

Let me expand on this general theme by developing in more detail the idea of stronger versus weaker levels of distinctions. The technical notation is refined in the following paragraphs

This research was supported by National Science Foundation Grant 8710163 from the Memory and Cognitive Processes Section. I am grateful to Robin Thomas and Trisha Van Zandt for help in preparation of the figures, proofreading, and some theoretical calculations, and to Kilsoon Cummings and Katherine Fuller for typing the manuscript. Helpful comments were offered on an earlier version of the manuscript by Edna Loehman (Department of Agricultural Economics, Purdue University), David Moore (Department of Statistics, Purdue University), and Peter Schönemann, Hans Colonius, and Trisha Van Zandt (Department of Psychology, Purdue University). Two referees provided useful critiques. Certain of the present results were presented at the nineteenth annual meeting of the Society for Mathematical Psychology at Harvard University, Cambridge, Massachusetts, August 1986.

Correspondence concerning this article should be addressed to James T. Townsend, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

and in the Appendix, but for now simply note that most of the discussion refers to the population or ideal statistical concepts. Thus, the term *distribution function* is used for the population cumulative frequency function and *density* is used for the population frequency function (which, for convenience and practicality, will be taken as being continuous in almost all cases). When necessary, "empirical" or "estimated" distributions or densities can be referred to as sampled approximations to the population functions, but for the most part the "sampling" questions must be put aside.

Consider again the administration of the aforementioned psychological test to the two groups. Suppose that the investigator entertains serious doubts about the satisfaction of the assumptions of the normal (Gaussian) distribution by the underlying population distributions. In that case, the finding of a difference in the means of the two groups could be quite weak in terms of the dominance hierarchy. However, suppose it were found that for any particular test score, there were always more "normals" who scored at that score or lower: That would be a more powerful finding than the one regarding a difference in the means. The more powerful result is equivalent to the distribution function of the normal group always being greater than that of the presumed pathological group. This ordering of the distribution functions in fact implies that the mean of the normal group is less than that of the hypothetically pathological group. The implication here is invariant under monotonic transformations of the scores. Thus, if one is certain within statistical error that the distribution functions are ordered, then in principle there is no need to test the means. However, the means could be different without the distribution functions being ordered. Thus the mean of the normals might be less than that of the pathological group, but for one or more particular scores, it might be that more of the pathological subjects achieved a score that was less than or equal to that particular score than was the case for the normal group. Hence, the distinction between the two groups is less strong in this case.

It is true that certain types of distribution functions such as the normal distribution, with equal standard deviations, have the property that if there is a difference between the two means, then the distribution functions will also be ordered. However, it is rare in the behavioral sciences for the researchers to possess a high degree of confidence that all the assumptions of normality are met in the underlying population distribution (see, e.g., Kendall & Stuart, 1973, Vol. 2, chap. 31; Mosteller & Tukey, 1977, chap. 1). Furthermore, I show how knowledge of this and other similar properties can aid theorists in their investigations. I discuss the ramifications with regard to this question as well as that relating to the exact strength of the measurement scale later in the article.

Consider two other examples of how the type of results reported here might be of use in psychological investigations. In the first example, suppose a researcher has developed a mathematical model of how some response variable will change as a function of an experimental variable. That model should then predict a distributional distinction at some level as the experimental variable is manipulated. Thus the researcher will likely wish to examine the data from the strongest level possible given the model. If it turns out that the distributions differ only at a weaker level, then this might call for a relaxation of certain aspects of the model. These aspects might be of a purely technical nature, or they may be more integrally related to the psychological processes under investigation.

In the second example, which is in a sense an extension of the first, I consider the investigation of mental architecture in a behavioral experiment. By architecture I mean the way in which various psychological subprocesses may be connected together. The simplest versions of a class of such architectures are serial (i.e., the subprocesses are arranged sequentially) or parallel (i.e., the subprocesses are arranged so that they operate simultaneously). More complex architectures can involve process interactions of considerable sophistication (Schweickert, 1978). It is possible to identify the type of mental architecture that is active in a particular cognitive task, if certain experimental variables affect specific subprocesses (Townsend & Schweickert, 1989; Schweickert & Townsend, 1989). In most, if not all psychological tasks, the subprocesses must be assumed to be stochastic; that is, they take a random duration to operate from trial to trial. The question thus arises as to what aspect of the subprocesses' distribution an experimental variable may affect. The result is that if the experimental variable affects that distribution at a level of moderate strength within the dominance hierarchy, then even rather complex architectures can be identified experimentally. These examples are amplified after the taxonomy is developed in more detail.

Before proceeding with that goal, I elaborate on certain features of the theoretical results. First, as was noted earlier, the scales of the underlying variate need only lie on an ordinal scale, although of course stronger scales are totally acceptable, as might be expected in such a situation. This may seem paradoxical because some results involve the arithmetic mean, and it is well known that many statements covering the mean are meaningful in a measurement sense only if the scale is at least interval. The apparent paradox is explained and my perspective on the measurement background given in the section *Level of Measurement Scale.*[1]

Next, all of the relations between two distributions that will be studied, with the exception of those pertaining to the normal equal variance case, are distribution-free. They are also parameter free except in the trivial sense that the mean and median are thought of as parameters. Thus, except as noted, no particular distribution or parameterization need be assumed to use these results. For some relations, statistical procedures are known and available in standard sources. Cases in which these have apparently not been developed are pointed out. Perhaps research on these issues will be stimulated.

The presentation in this article is limited to single variates, for instance, scores, reaction times, and the like. However, it may be possible to generalize the concepts to multivariate situations by such devices as defining vector $\mathbf{V}$ to be greater than vector $\mathbf{W}$ if and only if each entry of $\mathbf{V}$ is greater than the

[1] Roberts and Marcus-Roberts (1987) observed that in certain situations, as when a variable can attain only two possible values, even the ordering of the means is invariant across monotonic transformations. They also noted the fact established by Lehmann (1955) that an ordering of distribution functions implies an ordering of means. This relation is discussed later.

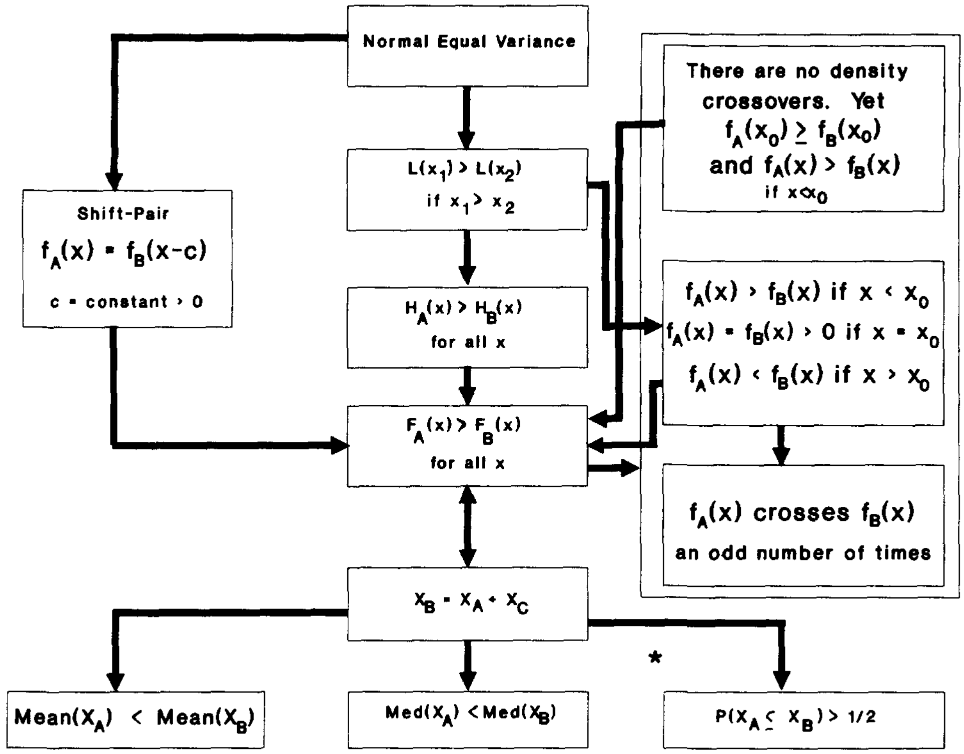

Figure 1. Implication graph showing the dominance relations of the various stochastic orders. (The arrows exhibit the implications. If no arrows connect two boxes, then there is no logical relation between those boxes. If the arrow goes both ways, then the order relations in the two boxes are logically equivalent. The arrow pointing from $F_A(x) > F_B(x)$ to the large box at the right indicates that one of the relations contained in the large box must hold. The meaning of $X_B = X_A + X_C$ is that there exists a random variable X_C such that $P(X_C \geq 0) = 1$ and $X_A + X_C$ has the same distribution as X_B. The asterisk at the lower right indicates that the implication holds if X_A and X_B are independent.)

corresponding entry of W. Of course, new problems may well be encountered there that do not arise in the simpler univariate case.

Hierarchy of Distinctions and Relevant Concepts

Brief Survey of Concepts

Most of the concepts in the theory will probably be very familiar to the reader, but there may be one or two that need introduction. This is done all at once here for compactness. Figure 1 gives the implication graph that is the focus of this discussion. The reader who is familiar with the various concepts found in Figure 1 should jump ahead to the next section.

There are a number of ways of specifying a probability distribution uniquely, any one of which can be derived from any of the others. For convenience and clarity, assume in the following that the measurements are on a continuum in the sense that the population frequency function is defined by a continuous density function on the line of real numbers for a single variate, or in the plane in the context of a bivariate measurement. That is, assume the general existence of a bivariate joint density $f_{AB}(x_A, x_B)$ in the measurements or "treatment" random variables x_A and x_B. Occasionally, stronger conditions are required in the proofs given in the Appendix.

Of principal interest are the relations of the marginal random variables as specified by the marginal densities $f_A(x_A)$ and $f_B(x_B)$. Of course when X_A and X_B are independent, corresponding to independent measurement or sampling, then $f_{AB}(x_A, x_B) = [f_A(x_A)][f_B(x_B)]$; that is, the joint density equals the product of the marginal densities.

Two unique ways of specifying a probability law are by way of its cumulative distribution functions or, more succinctly, "distribution function," and by way of its hazard function. In the case of single variates of most concern here, the distribution function $F(x)$ is the integral from $-\infty$ to x of the density function. Conversely, the density $f(x)$ can be viewed as the derivative of the distribution function.

The hazard function is given by dividing the density function by 1 minus the distribution function: $H(x) = f(x)/[1 - F(x)]$. The latter, $1 - F(x)$, is known as the survivor function, which is designated as $\bar{F}(x) = 1 - F(x)$. Both terms *hazard* and *survivor* come from applications in actuarial statistics, operations research, and quality control engineering and are not particularly appropriate for psychology, but they have now become conventional terms in psychological modeling with stochastic processes. An intuition of the hazard function is that if it is known that a randomly drawn score is at least x, then the conditional probability that it will in fact be just a little bigger than x—say, $x + c$, where c is very small—equals $H(x)c$. In the context of temporal phenomena, such as reaction time, it can be used to represent the probability that processing will be completed in the next instant, given that it has not been completed up until that time. The utility of the hazard function for actuarial statistics is evident. Life insurance companies are necessarily interested in the probability that people born in a certain year will die today (this year, etc.), given that they have survived until now. Similarly, quality control and maintenance engineers are concerned with the probability that a product will fail presently, conditioned on the event that it has lasted up until the present.

Other characteristics of probability laws, such as means and medians, have proven indispensable in statistics, although as will be shown, some of them convey much less information than others. Another important characteristic is the likelihood ratio, which simply is the ratio for any score x of the two density functions under consideration. Let $f_A(x)$ be the density for Treatment A and $f_B(x)$ be the density for Treatment B; then $L(x) = f_A(x)/f_B(x)$ is the likelihood ratio as x ranges over its possible values.

Hierarchy of Implication Graph

The dominance hierarchy of Figure 1 may be examined stepwise starting at the bottom with the mean and the median. All the propositions relating to Figure 1 are proven, or proofs are cited in the Appendix. Most of the relations hold even if the measurements X_A and X_B are dependent; exceptions are noted.

Mean versus median. Beginning at the bottom of Figure 1, the mean and median are seen under the assumption that Treatment B exceeds Treatment A, in the sense that the means or medians are also ordered in that fashion. Interestingly, neither the mean nor the median bears implications for the other in that either may be ordered in the manner shown and at the same time the other may have the reverse order. It is natural that these hold no implications for one another because it is well known that they capture somewhat different aspects of the parent probability distribution.[2]

Figure 2a shows two theoretical densities that can be used to illustrate this and several other aspects of Figure 1. Figure 2b exhibits two empirically estimated densities from an experiment on short-term memory search. Proposition 1 in the Appendix covers this case.

$P(X_A \leq X_B)$. The function $P(X_A \leq X_B)$ represents the probability that the X_A measurement is less than or equal to the X_B measurement. Thus, $P(X_A \leq X_B) > \frac{1}{2}$ says that Treatment B is at least as large as Treatment A more than one half of the time. It is rather surprising that this statement is probabilistically so weak, as is indicated in Figure 1. If X_A and X_B are independent, then this ordering is implied by the higher entries in Figure 1, and with or without independence it does not even force the means or medians to be ordered. Because ordering of the means or medians also does not force $P(X_A \leq X_B) > \frac{1}{2}$, it follows that all of these are logically unrelated to one another, and all are weaker than the other distinctions in the implication graph. The one exception is that if X_A and X_B are dependent, then it is possible to have $F_A(x) > F_B(x)$ yet $P(X_A \leq X_B) \leq \frac{1}{2}$, thus violating the dominance structure depicted in Figure 1. Proposition 2A treats the means and Proposition 2B the medians (see the Appendix) with respect to the present type of stochastic ordering. The theoretical example shown in Figure 2a yields $P(X_A \leq X_B) = .75$ to the second decimal. The empirical example of Figure 2b has $P(X_A \leq X_B) = .82$, approximately.

Adding a positive random variable and ordered distribution functions. The next entries in Figure 1 carry more force. One way to think about magnifying a set of measurements is to add a new positive random variable to the random variables representing the original scores. Hence, one could produce a "larger" population, described by the random variable X_B, by adding a positive random variable X_C to the "smaller" population described by X_A; that is, set $X_B = X_A + X_C$. A conceptually distinct but, as it turns out, logically equivalent dominance relation is the ordering of the two distribution functions $F_A(x) > F_B(x)$ for all x (except possibly where $F_A(x)$ and $F_B(x)$ both equal 0 or 1). As is noted in the introduction of this article, this is interpreted as meaning that for any score or measurement x, the frequency of those in Treatment A achieving x or lower is always larger than those in Treatment B achieving x or lower. Proposition 3 states that the two present dominance relations are equivalent in the sense that for any given distributional ordering, $F_A(x) > F_B(x)$, there exists a positive random variable X_C that can be added to the random variable for Treatment A, X_A, that will yield the distribution function of Treatment B and conversely. That is, $X_B = X_A + X_C$. The random variable, X_C, may not be independent of the random variable of Treatment A. Actually, for technical reasons, it is best to define a random variable X_A^* with a distribution identical to that of X_A and then increment X_A^*. The details are covered in Proposition 3 in the Appendix.

In any event, a distributional ordering of two treatments is for practical purposes the same as adding a possibly correlated random variable to the larger of the two distributions (i.e., the one associated with the smaller scores).

Proposition 4A (see the Appendix) uses the $F_A(x) > F_B(x)$ version to imply the ordering of the means and medians and with the added assumption of independence also implies that $P(X_A \leq X_B) > \frac{1}{2}$. The reverse implications do not hold, so $F_A(x) > F_B(x)$ and the equivalent $X_B = X_A + X_C$ are truly stronger.

Proposition 4B indicates that without independence, the dominance of $[F_A(x) > F_B(x)]$ over $[P(X_A \leq X_B) > \frac{1}{2}]$ may fail. An example in the Appendix confirms that one may find that Runner A is faster than Runner B in that for every time t, the probability that A beats that time is greater than the probability that B beats that time (i.e., $F_A(t) > F_B(t)$). Yet paradoxically, it may be that Runner B defeats Runner A in over one half of the races they run. The intuition for this and many other examples is that in a majority of cases, A and B run (say) at about the same speed and A always wins those races, leading to $P(X_A \leq X_B) < \frac{1}{2}$. However, in some races, B is very fast, while A is very slow, leading to $F_A(x) < F_B(x)$.

Figure 3a gives the distribution functions that correspond to the Figure 2a densities, whereas Figure 3b shows the empirical

[2] Another interesting sidelight is that there are situations, as with high-tailed distributions, in which the median may estimate the mean better than the arithmetic average of the data (e.g., Mosteller & Tukey, 1977, chap. 14).

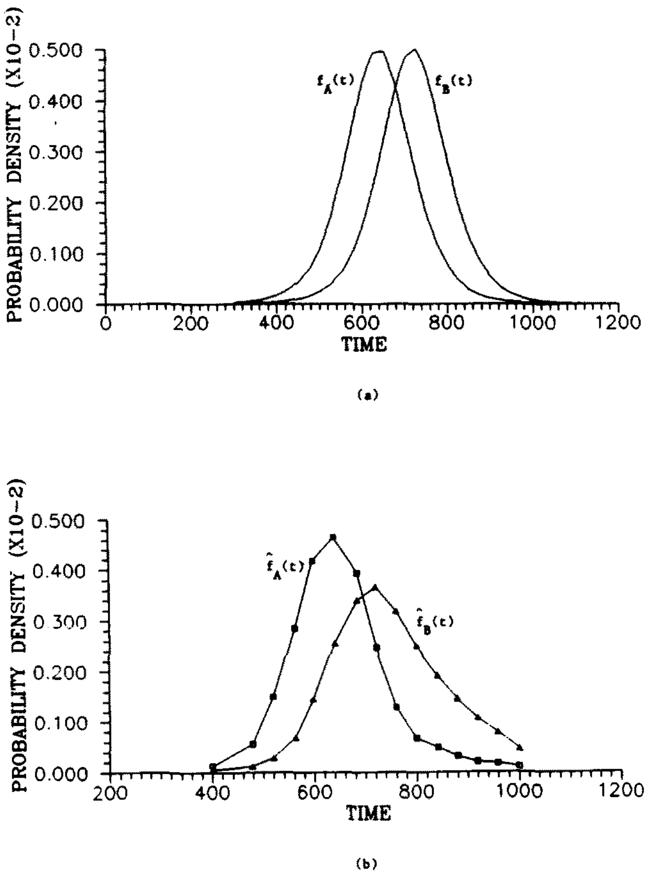

Figure 2. (a) Two population densities that are both logistic with the same variance and different means; (b) two estimated densities replotted from data of Ashby (1982). (Ashby estimated them from a study by Townsend & Roos, 1973, on short-term memory search. The term $\hat{f}_A(x)$ is the estimated reaction time density when the number of items in memory was 1, and $\hat{f}_B(x)$ is that for five items in memory.)

distribution functions corresponding to Figure 2b. The "ideal" functions of Figure 3a indicate that they are ordered, and the sample functions of Figure 3b suggest that the data from which they are derived satisfy distributional ordering.

Density function crossings. The question arises as to how the respective density functions relate to the ordering of their distributions. One interesting relation is found by studying the ways in which the two density functions cross over one another. Figure 1 indicates that the distributional ordering, $F_A(x) > F_B(x)$, implies that the density functions either must not cross at all or must cross an odd number of times. Conversely, if the number of crossings is 0 and the A density "gets started" before the B density, then $F_A(x) > F_B(x)$. A special case of the latter occurs when all of the A density lies to the left of the B density, an example of which was mentioned in the introduction of this article. Furthermore, if the densities cross exactly one time, the distribution functions are also ordered. However, an odd number of crossings greater than 1 is compatible with a distributional ordering but does not imply it. An immediate corollary is that if $f_A(x)$ and $f_B(x)$ cross each other an even number of times greater than 1, then no distributional ordering is possible, and $F_A(x)$ and $F_B(x)$ must cross.

Observe that because an odd number of crossings of the frequency functions might be accompanied by a distributional ordering, a single crossing of the frequency functions is stronger than the distributional ordering and can be looked for in the data with this in mind. Propositions 5 and 6 (see the Appendix) are relevant to these statements. The ideal densities shown in Figure 2a satisfy the single crossover point criterion, and the empirical cases in Figure 2b appear to do so also.

Shifting a distribution. A distinction that is definitely stronger than that of the ordering of distribution functions is that of $F_B(x) = F_A(x - c)$, where c is some positive constant. This immediately implies that $F_A(x) > F_B(x)$. That is, the distribution functions are ordered if one is obtained from the other by a shift of the measured variable. It is fairly evident, as shown by Proposition 7 (Appendix), that distribution functions can be ordered without the strong assumption of shifting, so shifting is stronger and more constraining than simple distribution ordering. The concept of creating a family of distributions by taking an original distribution and performing arbitrary shifts on it is important in mathematical statistics.

Note that in Proposition 3, if the distribution on X_C is concentrated, with all probability on one value $X_c = C$, we have the shift family. Interestingly, in the present case, the representation (by shift) is unique, whereas when $F_C(x_c)$ is nonnull, the repre-

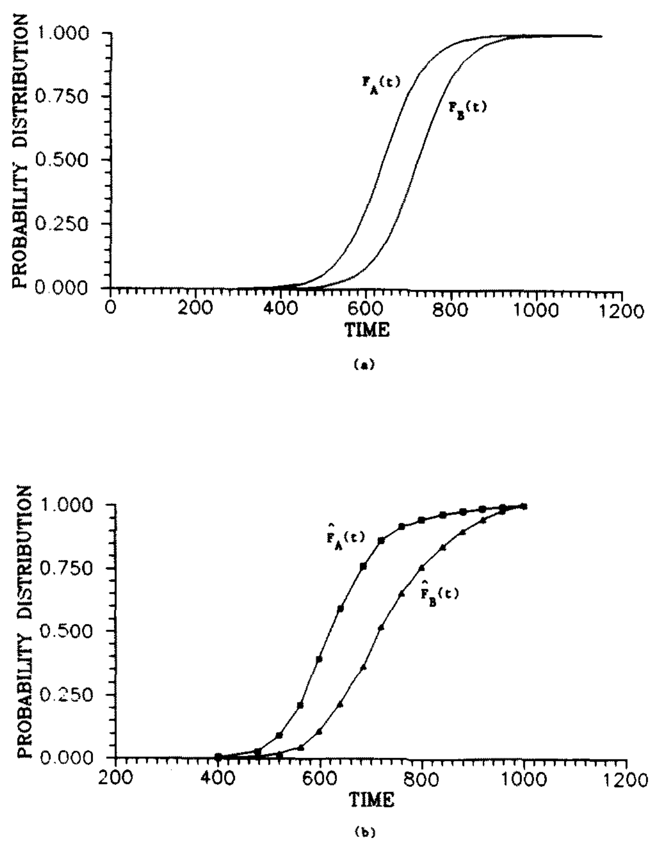

Figure 3. (a) $F_A(X)$ and $F_B(x)$ are the cumulative distribution functions associated with $f_A(x)$ and $f_B(x)$ of Figure 2a; (b) $\hat{F}_A(x)$ and $\hat{F}_B(x)$ are the empirical distribution functions associated with the estimated densities of Figure 2b.

senting additive random variable (and joint distribution) is in general not unique.

For simplicity, the relation of shift pairs to density crossings has been omitted in Figure 1. Basically, matters are inelegant in the most general cases, in that weird shift-pair densities may cross an odd or even number of times. However, if f is unimodal by way of its derivative being positive for x less than the mode and negative for x greater than the mode, then the two shift densities cross exactly once. This is a strong manner of getting the distribution function ordering, which is shown in Figure 1. Note that $f_B(x)$ in Figure 2a is a shift of $f_A(x)$. The empirical densities of Figure 2b clearly do not satisfy this criterion.

Hazard functions. The next distinction is based on an ordering for all x of the respective hazard functions. The first result here is that this ordering implies an ordering of the associated distribution functions but not conversely, so that the ordering of hazard functions is stronger than the ordering of distribution functions. This distinction, like the others, can be useful regardless of the particular area of application. However, some may be especially cogent in specific contexts. For instance, hazard functions carry considerable intuitive appeal when used in a time-dependent situation. When x represents time and $H_A(x) > H_B(x)$, the interpretation is that the instantaneous probability of finishing a task, given that it has not been completed up until time x, is always greater for Treatment A than for Treatment B. Because this distinction is stronger than an ordering of distribution functions, it follows that for all times x, the probability that a Treatment A subject (a particular subprocess, etc.) is done by time x, is always greater than that a Treatment B subject is done by that time. Proposition 8 cites a proof of this finding (see the Appendix). It is important to note that an ordering of two hazard functions does not imply that either is itself a monotonic function of x. Figure 4a shows the ordered hazard functions associated with $f_A(x)$ and $f_B(x)$ of Figure 2a. Figure 4b illustrates empirical hazard functions estimated in a rough way from the densities of Figure 2b. Proposition 9 (Appendix) shows that a shift pair of distributions need not force their hazard functions to be ordered, or vice versa. Proposition 10 demonstrates that ordered hazard functions do not imply a single density crossover point (see the Appendix). Of course, if there is at least one crossover point, there there is an odd number of crossovers, because otherwise the associated distribution functions would not be ordered.

Likelihood ratio. The next distinction, the likelihood ratio, is analyzed because of its importance in statistical decision theory. It was surprising to learn that the assumption that this function, defined as $f_A(x)/f_B(x)$, is monotonically decreasing is even more powerful than the ordering on hazard functions. The former implies the latter, but not vice versa as is shown in Proposition 11 (see the Appendix). Intuitively, a monotonically decreasing likelihood ratio is saying that as one moves along the axis of measurement, the evidence is always shifting continuously from Treatment A toward that of Treatment B. An example of its use in psychology is its application to signal detection theory (e.g., Green & Swets, 1966). There, as the decision criterion, figured as the likelihood ratio (equivalent to likelihood function here), moves from minus infinity to plus infinity, it

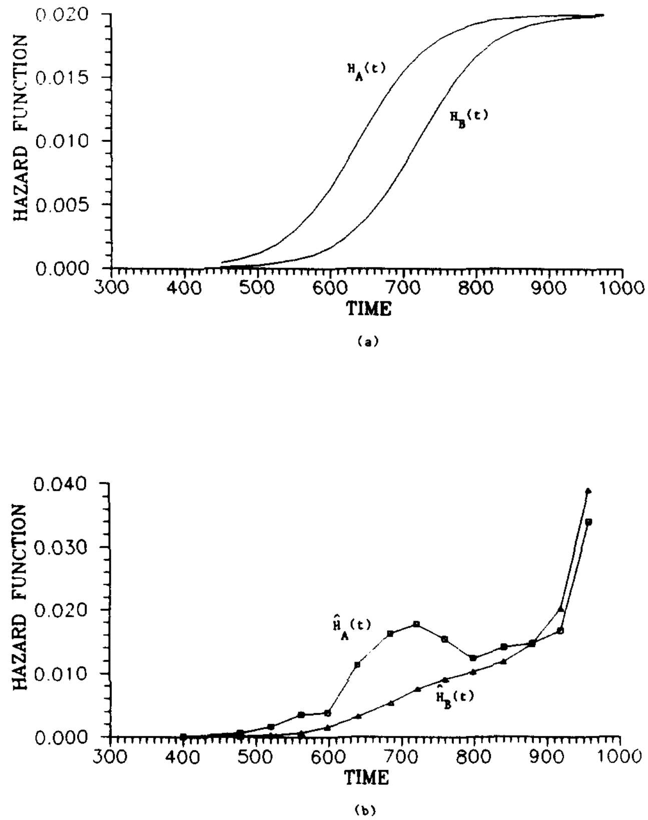

Figure 4. (a) Hazard functions correspond to the probability distributions from Figures 2a and 3a; (b) a rough estimate of the hazard functions corresponding to the empirical distributions of Figures 2b and 3b. (It was not feasible to use the more sophisticated techniques of Bloxom, 1983.)

monotonically decreases (that theory typically uses the reciprocal of my function so that Green and Swets' function increases where the present function decreases) and in the process sweeps out the well-known receiver operator characteristic (ROC) curve. However, it is not typically used in that theory to distinguish between distributions per se.

Proposition 12 suggests that for most usual cases, strictly decreasing likelihood ratios imply a single density crossover point. This proposition is stated loosely but appears to suffice for present purposes. Proposition 13 gives an example showing that a strictly decreasing likelihood ratio does not imply that the A and B densities are related by a shift.

Figure 5a demonstrates that $L(x)$ for the example of Figure 2a is monotonic in the proper manner. Figure 5b shows the likelihood ratio generated by the empirical densities of Figure 2b. It is intriguing that at least for the sizeable differences in processing load in Figures 2b, 3b, 4b, and 5b (one item vs. five items in short-term memory) even the strongest distinction of monotonic likelihood ratio appears to hold. The hazard functions seem to lose their ordering for time greater than 900 ms, but that is in the extreme tail of the $\hat{f}_A$ density. Certainly, caution must be used here owing to the rough methods of estimation. Nevertheless, the example offers hope that at least in some contexts, even the stronger distinctions may not be chimerical.

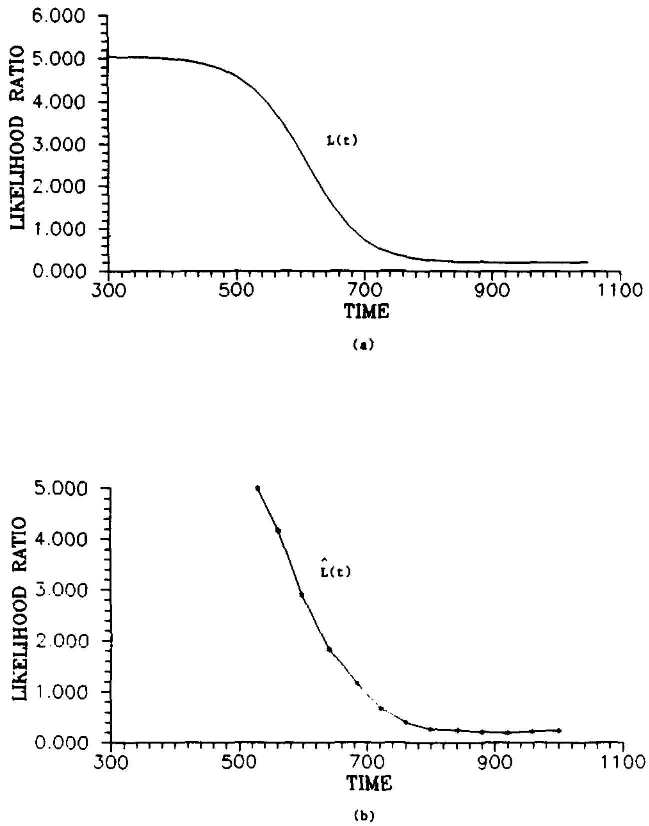

Figure 5. (a) $L(t)$ is the likelihood function composed of the densities shown in Figure 2a; (b) $\hat{L}(t)$ is an estimate of the likelihood function taken from the estimated densities in Figure 2b.

Shift family of normal distributions. The strongest distinction considered in this article is that of the shift family of normal distributions; that is, the family of normal distributions for which the means may vary but the variances are the same or homogeneous. As was previously noted, this obviously dominates the general shift family in the spirit of the hierarchy. Moreover, along with the properties of the normal distribution family, all the other distinctions are implied. Thus, the likelihood ratio is monotonically decreasing, the hazard functions and distribution functions are ordered, one distribution function is derived from the other by adding a constant to the first random variable (a special case of adding a new positive random variable to the first random variable), the one-point crossing property of densities holds, and so on. The proof of the strictly monotonic likelihood ratio is given in Proposition 14. The fact that $f_A(x)$ and $f_B(x)$ are related by a shift is obvious; observe that the densities of Figure 2a are not normal, although they provide useful approximations to normal distributions. The empirical densities of Figure 1b are decidedly not normal, as is the case with most reaction time distributions. Observe that, like the ordering on means and shift pairs, the property of normality is not in general preserved under monotonic transformations.

A brief consideration of measurement and statistical issues relating to the hierarchy follows.

Implications and Issues Concerning Measurement and Statistics

Level of Measurement Scale

In this and the next section the implication graph in relation to the traditional question of discriminating between two treatment groups, conditions, or similar manipulations is discussed. The issue of the scale of measurement is intimately related to the discrimination question. With provisos to be stated later, the properties of the implication graph hold regardless of the scale of measurement except the nominal scale, so that it is always true that the higher up in the graph that a relation holds, the stronger the conclusions. Nevertheless, the ordinal properties of the tree are of particular interest for psychology because of the difficulty in establishing scales at the interval or ratio level. The level of measurement issue is briefly dealt with, then the general problem of using the various orderings of the tree in statistical tests is discussed.

There is a long history of controversy about whether the strength of the measurement—that is, nominal, ordinal, interval, or ratio—should affect the statistical operations that are visited on one's data and on the investigator's conclusions. This controversy is not the main focus of the present research. Some recent debates on the subject, which also include references to other contemporary and older works, are Davison and Sharma (1988), Gaito (1980), Michell (1986), Townsend and Ashby (1984). This article follows the foundational, or what Michell (1986) refers to as the representational, approach (e.g., Krantz, Luce, Suppes & Tversky, 1971; Narens, 1985; Pfanzagl, 1968; Stevens, 1951; Roberts, 1979; Suppes & Zinnes, 1963). This is not the place for detail on a highly technical subject, but basically this approach views a statistical statement as meaningful if and only if the truth of the statement is invariant over admissible transformations of the scale.[3] In the case of ordinal scales, the class of admissible transformations is that of strictly monotonic functions.

It should be noted that there are a number of technical conditions that must be satisfied to permit a qualitative probability distribution to be defined on a relational system (see, e.g., Fishburn & Roberts, 1989; Krantz et al., 1971, chap. 5; Van Lier, 1989). It is the relational system that captures the regularity in empirical data that permit the mapping into a numerical representation. In fact, in most experiments, which naturally contain only a finite number of points, there is no definite way to assess even the requirements of continuity of the scale (Pfanzagl, 1968, pp. 76–79). Such conundrums appear to exist in all sciences, not just psychology. As in most cases of application of measurement and probability, it is simply assumed here that the underlying technical conditions are met, so that the appro-

[3] Actually, matters can get rather Byzantine in the most general cases (e.g., Roberts & Franke, 1976). However, suppose the scales are regular in the sense that if f and g are two scales for the same relational system, then there is a transformation ϕ such that $g = \phi(f)$. Then the notion of admissible transformation seems to adequately capture the idea of meaningfulness (see also Falmagne & Narens, 1983; Roberts, 1979, 1985).

priate scales and probability distributions on these scales are legitimate. In this context, the reader may interpret "ordinal scale" to mean "ordinal scale with qualitative probability". Ancillary remarks are made on this subject in the Appendix.

One useful viewpoint in many psychological contexts is in terms of a measured manifest variable that is itself a function of a latent variable, the true object of interest to the psychologist. Now it is common in statistical inference to use (i.e., make statements about) means in ways that demand interval level measurement. For instance, most parametric tests such as Student's t test, multiple regression, and analysis of variance (ANOVA), which likely make up the bulk of psychological applications, rely on normal distributions. Linear (or affine) transformations of normally distributed variables are again normal. Also, if two distributions have different means but the same variance, then the ratio of the difference in means to the common standard deviation remains invariant under affine transformations (i.e., $y = ax + b$, where a and b are constants).

These two properties cannot be overemphasized in terms of their importance for statistics. But the catch is that only measurements at the interval or ratio level are restricted to affine or linear transformations ($y = ax$; i.e., $b = 0$ in the linear, special case of the affine representation). In particular, ordinal scale measurement permits any monotonic transformation, and this will enable any distribution, in particular the normal distribution, to be perturbed to some other, often unspecified or unstudied, distribution. Thus, the manifest variable must be an affine function of the latent variable for the foregoing techniques to be meaningful in most circumstances. If it is only ordinally related to the latent variable, it would seem nonsensical to perform such tests. Furthermore, even if the manifest and latent variables were one and the same, any strict monotonic transformation is still admissible.

In fact, it does seem likely that many psychological latent variables lie at best on ordinal scales and thus carry only order information. As such, they may legitimately undergo any monotonic transformation, which to the adherents of the foundational (representational) view, effectively precludes the usual battery of parametric analyses.[4]

Most of the orderings depicted in Figure 1 are invariant under monotonic transformations. In other words, the class of ordered hazard functions is closed under that class of transformations; so is the class of ordered distribution functions, and so on. However, as pointed out earlier, means do not necessarily stay ordered; furthermore, the shift family of distribution functions is not closed under monotonic transformations, although any such pair does remain ordered. Of course, the normal distribution family is also not closed under monotonic transformations.

Furthermore, some care must be taken with the entries relating the distribution function ordering, $F_A(x) > F_B(x)$ and $X_B = X_A + X_C$. Recall that the distribution function ordering is logically equivalent to adding a positive random variable to X_A. The distribution function ordering is invariant under monotonic transformations of X_A, X_B but of course, the distribution of X_C is expected to change accordingly. Conversely, if X_A and X_C are altered monotonically, then the resulting X_B will be different than before the transformation was exacted. However, none of the logical implications in Figure 1 are harmed by these facts.

Within these guidelines, the implications of Figure 1 are not altered by monotonic transformations, although as just noted, the particular entities satisfying a particular statement or relation may change in the cases of two normally distributed variables with equal variance, a shift pair and a pair of ordered means. Overall then, and with these caveats, as long as the latent variable and the manifest variable are monotonically related, the relations of Figure 1 continue to hold if they held before the transformation and fail to hold if they violated the relation before the transformation.

Davison and Sharma (1988) have pointed out that if the densities cross exactly once, then ordered means stay ordered. Thus, if the latent and manifest variables are monotonically related, then the means of both must be ordered in the same way. The dominance hierarchy of Figure 1 shows that the single crossover condition is even stronger, in that the distribution functions themselves stay ordered in going from the manifest to the latent variable. The latter in turn produces the mean ordering developed by Davison and Sharma, as is shown in Figure 1.

Davison and Sharma (1988) also noted that the example used by Townsend and Ashby (1984) does not satisfy the single crossover condition, and hence the order of the means need not be preserved under a monotonic transformation. Actually, in that example the difference in the means in relation to the appropriate variance was enlarged from 0 to an arbitrary number, hence an arbitrary level of significance, but their basic point remains valid. From a purist's viewpoint, it is in any case meaningless to speak of "statistical significance" when the normal distribution is only one of an infinite number of distributions that applies to this situation. Also, in general, two arbitrary distributions cannot be changed to both be normal through a single transformation, so that one may not somehow refer to a normal "canonical" form for two such distributions in order to provide for a statistical test.

From the opposite direction, what if a set of measurements is on an interval or ratio scale? For instance, when used in its physical sense rather than as a measure of a psychological variable, reaction time may lie on a ratio scale (cf. Krantz, 1972; Micko, 1969; Townsend, in press; Townsend & Ashby, 1983, pp. 387–390).[5] As has been noted, reaction time data are almost

---

[4] As was rightly pointed out by a referee, an analysis may not be ruled out simply because it is parametric. Suppose that a statement, for instance one involving a comparison of two distributions, is based on a parameter of the assumed family of distributions and that the truth of this statement is invariant under all strictly monotonic transformations. Then it follows that one has a legitimate ordinal comparison that nevertheless is "parametric." However, it is a fact that most commonplace statistical tests do not satisfy the foregoing condition but rather demand at least interval level measurement.

[5] A referee pointed out that not all investigators believe that reaction time lies on a ratio scale and cited Krantz (1972) in this regard. This issue obviously cannot be dealt with in detail here. Elsewhere, I argue that processing time consumed by a psychological mechanism (which ultimately refers to action by a congeries of neural elements) can in certain cases be treated exactly as if one were measuring the fall-time of an apple from a tree (Townsend, in press). The reaction time scale would then be ratio in such contexts. On the other hand, if reaction time, or virtually any other dependent variable, is taken as a measure of a psychological entity, such as response strength in Krantz (1972),

never normal in appearance, but rather positively skewed. If the scale is ratio, it is illegitimate to transform the scale or data in order to render the distribution more normal in form. Thus, the investigator must develop parametric tests that are specific to a family of distributions that seem to match the sample frequency functions, or they must use distribution-free tests (see the next section). The latter may be in some cases the better and more efficient approach.

Statistical Analyses

Nonparametric statistics treats estimation and hypothesis testing in such a way that the specific parameters of a distribution may be safely ignored. Distribution-free statistics relaxes things further in that even the form of the distribution can be neglected. Thus, *distribution-free* implies parameter-free, but not conversely. As was noted earlier, except insofar as, say, the mean is considered to be a parameter of a distribution, the present methodology is both distribution- and parameter-free. By convention, nonparametric or parametric-free tests are often also distribution-free. (But see Kendall & Stuart, 1973, Vol. 2, for a discussion of this issue.)

Presently, there seems to be no theoretical structure existing in the field of statistical hypothesis testing corresponding to the present hierarchy as a whole. The implication graph of Figure 1 originally grew out of questions related to the strength of cognitive capacity limitations as revealed through reaction time statistics (e.g., Townsend, 1974; Townsend & Ashby, 1978). That is, the mental work load is varied, and some statistic or function is used to assess the effects on reaction time. All of the propositions relating to Figure 1 (see the Appendix) have been proven independently by myself and my associates. However, we subsequently uncovered lines of research in other domains in which stochastic orders have proven of importance. Early related papers are Rubin (1951) and Lehmann (1955). Stochastic orders are related to the concept of majorization, which provides a general approach to the study of probabilistic dominance relations (see, e.g., Marshall & Olkin, 1979). A readable beginning source for other approaches and applications outside psychology of some of the orders of Figure 1 is Ross (1983), which also includes additional references.

There are statistical procedures for some of the distinctions displayed in Figure 1, but several distinctions lack such methods. In addition, whereas some of the distinctions may be tested without recourse to estimation, in other cases, estimation procedures must be carried out before asking whether the distinction holds in the investigator's data. (This seems an apt place to remind the reader about the danger of α inflation that may accrue with a series of statistical tests.) Finally, there seems to be little known about the general relation of statistical power to the dominance hierarchy of relations, which is presented below. It is also important topic for further research.

Several of the relations of Figure 1 are illustrated by data from the experimental literature. It is convenient to begin this time at the top of Figure 1, where the strongest distinctions reside. The shift family of normal distributions (i.e., equal variance, but possible unequal means) is the cornerstone for much of parametric statistics, either directly, as in the z test or ANOVA, or because it enters into other theoretical formulae or derivations, as in the t test or central chi-square distribution.[6] If the model itself is true, then all the major distinctions of Figure 1 are in force (see the previous section for caveats).

The left side of Figure 1 shows $f_A(x)$ and $f_B(x)$ related by a shift transformation. Nonnormal shift families of distributions are popular in theoretical work in statistics, and it happily turns out that there is a test of the hypothesis that F_B is a shifted version of F_A. It appears in Pitman (1938) and is discussed by Kendall and Stuart (1973, Vol. 2, pp. 505–510).

Next, the likelihood ratio being used, appearing second down in the center of Figure 1, should not be confused with the likelihood function used in the so-called likelihood ratio test (e.g., Wilks, 1962, p. 402) that provides for a test of composite hypotheses by using a ratio of densities based on a null hypothesis versus a broader based hypothesis (e.g., testing a specific value of a mean vs. any positive mean). There are mathematical relations between the latter conception and mine, but these are beyond the present scope of this article. As far as is known, there are no available tests that directly assay whether the likelihood ratio is monotonically decreasing. Probably it would be necessary to estimate the density functions first before evaluating the performance of the function as the variate or scores increase. Even then, there appears to be no test available to statistically evaluate that question. Thus, at this time, this very powerful distinction apparently must be evaluated in a qualitative and statistically inexact fashion. One might use procedures to estimate the density functions in a distribution-free way (see the discussion on estimation at the end of this section) and then observe the results with an eye to looking for obvious or extreme violations of monotonicity of the likelihood function.

The state of the art with regard to hazard functions is also less than what one might hope, although here, too, progress has been made, at least with regard to estimation of the underlying hazard functions under, say, two conditions. However, again there appear to be no known procedures for statistically evaluating the order of the two functions. For that one requires either a distribution-free theory, corresponding to what exists in the case of distribution function ordering (see the following section), or a derivation of sampling distributions of the sample hazard functions, based on underlying parametric distributions. Nevertheless, there are early indications that empirical situations in psychology exist for which the hazard functions are indeed ordered; this will be seen in the next section.

The question of where and how many times two densities cross also has apparently not received great attention from statisticians. From a somewhat different viewpoint, the issue of where they cross has been shown to be of theoretical interest by Ashby (1982) and Ratcliff (1988) in cognitive and perceptual

then one must proceed to obey the tenets of foundational measurement in order to establish the scale type. Finally, it is not obvious that this assertion or its contrary is subject to logical proof. Perhaps arguing by analogy to theory and measurement in other sciences is as close as one can come.

[6] Recall that the central chi-square distribution may be derived as the sum of a set of normally distributed random variables, each with a mean equal to zero and a standard deviation equal to 1.

search experiments. Thus, it may be that interest can be provoked in several arenas that will encourage more attention to this problem.

Matters improve again with the distribution function ordering $F_A(x) > F_B(x)$. Here one can rely on the well-known Kolmogorov-Smirnov test (Kendall & Stuart, 1973, pp. 473–476; Siegel, 1956, pp. 127–136). The basic two-sided test simply decides whether two distribution functions are the same. If the null hypothesis of equality is rejected, then one may proceed to a one-sided test with the null hypothesis being that $F_A \geq F_B$. If the latter is accepted, one has the "predicted" ordering. If the one-sided null hypothesis is rejected, one infers that either a true crossover of the distribution functions exists or that the ordering goes in the opposite direction. The last possibility can be checked by another one-sided test.

There appear to be no statistical tests that address the question of whether a given distribution function can be represented as being generated by the random variable of another distribution plus an added positive random variable. However, as can be seen from Figure 1, the distribution function ordering logically implies the existence of such a random variable, and vice versa, so the appropriate tests for the former can also be used with the latter.

Because means can be reordered by monotonic transformations, there can be no ordinal tests involving them. However, there are a number of tests for the medians (see, e.g., Kendall & Stuart, 1973, Vol. 2, chap. 12, or Siegel, 1956, pp. 111–116). The statement that $P(X_A \leq X_B) > \frac{1}{2}$ forms a composite one-tail alternative to the null hypothesis that $P(X_A \leq X_B) = \frac{1}{2}$. This is readily assessed by the well-known sign test (e.g., Siegel, 1956, pp. 68–75).

Several of these orderings, such as number of density crossings and monotonicity of the likelihood function, may depend on adequate estimation of the densities or other functions, such as the hazard function. It may be surprising that although $f(x)$, $F(x)$, and $H(x)$ all yield total information about a distribution, a good estimate (e.g., optimal in some sense) of one of them does not necessarily lead to a good estimate of the others. For instance, the empirical distribution function proves to be a good estimate of the population distribution function yet does not immediately lead to an adequate estimate of the density function representation. The distribution function F has been most studied, and the situation for the Kolmogorov-Smirnov test is simplicity itself because the test is accomplished on the two usual cumulative frequency functions (also known as the *cumulative step function* and the *empirical distribution function* among other less well-known titles). This is easily obtained, and no difficult estimation problems intercede (for more detail see Wilks, 1962, pp. 454–459).

Matters are not so complete in the case of the density and hazard functions, but densities have occupied a good deal of attention from statisticians, engineers, and quantitative psychologists in recent years and hazard functions somewhat less often. Psychologists might begin with some recent papers of Bloxom (1983, 1984, 1985). Bloxom provided his own approach both to densities and hazard functions but also aided the reader in accessing the general literature on these topics. Furthermore, some of his techniques are especially important for the process theorist because they view a particular distribution as being a component in an overall processing system (as in serial, parallel, etc.).

Theoretical Applications

By *theoretical applications,* I refer to applications whose main intent is other than the immediate discrimination between two experimental treatments, in the broad sense. Thus, included in this section is the problem of discerning how alteration of a distribution, for instance, through change of a parameter, will affect theoretical predictions. For example, suppose it is postulated that an experimental manipulation affects a certain parameter of a distribution in a psychological model of the experimental task. That distribution, then, as part of the model, will produce a prediction that may fall into an ordering of the type pictured in Figure 1. In many such cases, one can then refer back to the previous section for points relating to the appropriate statistical tests and other pertinent information. There is another more subtle way in which the implication graph can be used that is discussed subsequently.

An example of a distribution function ordering was given by Townsend and Ashby (1983, pp. 210–211), who showed that independent serial processes and independent parallel processes both predict distribution function orderings as the number of items to process increases (see also Lupker & Theios, 1977; Sternberg, 1973; Vorberg, 1981). Therefore both serial and parallel models can predict the distribution ordering of reaction times found in a reanalysis of the Townsend and Roos (1973) study on memory and visual display search, referred to in an earlier section. Figure 3b is an example of the results. The reader is referred to Townsend and Ashby (1983, chap. 8) for related issues and tests.

Townsend and Ashby (1983, pp. 272–289) applied some of the results of the present article to a parallel horse race model based on two parallel counters (also called *accumulators*). Consider an application to a pattern classification experiment in which the subject must respond as to which of two classes a stimulus pattern belongs. Suppose that the correct counter always counts at a faster rate than the incorrect counter. Then if the criterion number of counts for the correct counter to "detect" is less than or equal to that for the incorrect counter, the conditional distribution function of the correct counter will always be greater than that for the incorrect (error) counter, say $F_c(t) > F_e(t)$. It immediately follows that the average correct response times must be faster than error response times. Furthermore, when both are running in parallel, the correct counter will beat the incorrect counter to a detect response more than one half of the time, corresponding to a probability correct greater than one half. The latter is, of course, a distribution-free prediction corresponding to $P(X_A \leq X_B) > \frac{1}{2}$ in Figure 1. This type of theoretical analysis already seems to be proving of value in testing competing models of two-choice discrimination (e.g., Smith & Vickers, 1988).

Busemeyer and his colleagues have used stochastic dominance orderings in obtaining predictions. They have found, as we have, that sometimes the stronger level is actually the easier to prove. Thus, in Busemeyer, Forsyth, and Nozawa's (1988) theoretical comparison of two dynamic choice models, it was shown that a likelihood function was monotonic in order to prove the ordering of two theoretical means. In another instance, Busemeyer and Rapoport (1988) argued from the empirical finding of a crossing of two distribution functions that a certain model that predicted that an ordering would occur must be wrong.

Burbeck and Luce (1982) were not directly testing treatment differences using the Figure 1 hierarchy, but some of their results are interesting from that perspective. Seven of the nine estimated hazard function pairs, one for the less intense tone and one for the more intense tone, appear to be ordered (Burbeck & Luce, 1982, Figure 5). From the implication graph in Figure 1, it can be seen that strong consequences ensue in terms of the speed of completion of the underlying detection process under the two conditions. It is also interesting that one of their plausible single process models, which turned out not to fit so well, clearly fails the hazard function ordering as a function of its shape parameter (see Burbeck & Luce, 1982, Figure 6).

In a study on pain tolerance, Stevenson, Kanfer, and Higgins (1984) manipulated goal and cue information, regarding the time remaining or time elapsed of the pain stimulus. Among other effects was the finding that the empirical distribution function of termination in the cue-to-cue interval was greater for the cued than for the uncued subjects. This result was interpreted to indicate that the cued subjects, given that they had not yet stopped at the point of a cue, tended to quit earlier in the interval before the next cue than did the control group. Thus, the cues may have served as psychological epoch markers as much as motivators to remain in pain longer overall.

The idea of stochastic ordering has played an important role in certain research areas of economics and decision making. Two starting references here are Hadar and Russell (1969) and Brumelle and Vickson (1975).

The final example extends the prediction concept. Schweickert and Townsend (Schweickert & Townsend, 1989; Townsend & Schweickert, 1985; Townsend & Schweickert, 1989) have developed a method of identifying mental architecture within a broad class of mental networks (directed acyclic networks endowed with probability distributions on the durations consumed by the subprocesses; see, e.g., Fisher & Goldstein, 1983; Schweickert, 1978). A key feature is that internal subprocesses are experimentally affected by way of incrementing their processing time random variables. This incrementing conforms to the addition of a positive random variable as is seen in Figure 1 and therefore is at the same level of strength as a distribution function ordering. When two subprocesses are so manipulated, it is possible to determine the type of architecture in which the two subprocesses are embedded, by observations at the level of overall mean reaction times.

Summary and Conclusions

A theory of hierarchical inference was developed and a graph that exhibits the implication relations among the consequent set of order relations has been presented in Figure 1. The implication graph shows that there is a hierarchy of strength holding among various individual indices of ordering. The form of the ranking of one treatment over another frequently was in terms of an ordering of two statistics or functions, although in certain cases, density crosses or shifts of a distribution created ordering relations. It is proposed that it is to the advantage of an experimenter to determine the strongest way in which one treatment group differs from another, for instance, as an ordering of the distribution functions instead of simply a difference in the means. The ordering relations in Figure 1 where statistical tests exist are pointed out as well as certain references to literature on associated estimation problems. It is not possible to develop or discuss sample and estimation themes in detail. I hope that the potential utility of these and similar dominance relations will encourage statistical theorists and psychologists to develop further hypothesis tests based on them.

For those who hold the level of measurement (ratio vs. interval vs. ordinal) to be important in the application of statistics, the invariance of the Figure 1 relations (a form of statistical statement) under monotonic transformations could be an important advantage, along with other distribution-free statistical procedures. It has been further shown that the relations of the implication graph could be helpful in theoretical investigations, particularly in assessing the level at which a model or theory would predict a difference between two treatments (as produced through manipulation of a theoretical distribution). Such results tell researchers at what level they should expect to detect experimental differences. Also, a difference on this dimension between two models may itself indicate a testable distinction or, other things being equal, suggest that a model predicting the stronger difference, in terms of Figure 1, is the preferred model because it is more falsifiable. Finally, a way of using the method to identify internal cognitive processing architecture has been discussed. Several examples of use of the concepts, some hypothetical and some from the psychological literature, have been offered. Most of them were from the perceptual and cognitive literature because that is the domain of this author. I hope that those from other areas will find the proposed theory of value.

References

Ashby, F. G. (1982). Testing the assumptions of exponential additive reaction time models. *Memory and Cognition, 10,* 125–134.

Bloxom, B. (1983). Some problems in estimating response time distributions. In H. Wainer & S. Messick (Eds.), *Principles of psychological measurement: A Festschrift in honor of Fredrick M. Lord.* Hillsdale, NJ: Erlbaum.

Bloxom, B. (1984). Estimating response time hazard functions: An exposition and an extension. *Journal of Mathematical Psychology, 28,* 401–420.

Bloxom, B. (1985). A constrained spline estimator of a hazard function. *Psychometrica, 50,* 301–321.

Brumelle, S. L., & Vickson, R. G. (1975). An unified approach to stochastic dominance. In W. Ziemba & R. G. Vickson (Eds.), *Stochastic optimization models in finance.* New York: Academic Press.

Burbeck, S. L., & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception and Psychophysics, 32,* 117–133.

Busemeyer, J. R., Forsyth, B., & Nozawa, G. (1988). Comparisons of elimination by aspects and suppression of aspects choice models based on choice response time. *Journal of Mathematical Psychology, 32,* 341–349.

Busemeyer, J. R., & Rapoport, A. (1988). Psychological theory of deferred decision making. *Journal of Mathematical Psychology, 32,* 91–135.

Chung, K. L. (1974). *A course in probability theory.* San Diego, CA: Academic Press.

Davison, M. L., & Sharma, A. R. (1988). Parametric statistics and levels of measurement. *Psychological Bulletin, 104,* 137–144.

Falmagne, J. C., & Narens, L. (1983). Scales and meaningfulness of quantitative laws. *Synthesis, 55,* 287–325.

Fishburn, P. C., & Roberts, F. S. (1989). Axioms for unique subjective probability on finite sets. *Journal of Mathematical Psychology, 33,* 117–130.

Fisher, D. L., & Goldstein, W. M. (1983). Stochastic PERT networks as models of cognition: Derivation of the mean, variance, and distribution of reaction time using order-of-processing space (OP) diagrams. *Journal of Mathematical Psychology, 27,* 121–151.

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin, 87,* 564–567.

Green D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hadar, J., & Russell, W. R. (1969). Rules for ordering uncertain prospects. *American Economic Reviews, 59,* 25–54.

Kendall, M. G., & Stuart, A. (1973). *The advanced theory of statistics* (Vol. 2). New York: Hafner.

Krantz, D. H. (1972). Measurement structure and psychological laws. *Science, 175,* 1427–1435.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.

Lehmann, E. L. (1955). Ordered families of distributions. *Annals of Mathematical Statistics, 26,* 399–419.

Lupker, S. J., & Theios, J. (1977). Further test of the two-state model for choice reaction times. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 496–504.

Marshall, A., & Olkin, I. (1979). *Inequalities: Theory of majorization and its applications.* New York: Academic Press.

Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin, 100,* 398–407.

Micko, H. C. (1969). A psychological scale for reaction time measurement. *Acta Psychologica, 30,* 324–335.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression.* Reading, MA: Addison-Wesley.

Narens, L. (1985). *Abstract measurement theory.* Cambridge, MA: MIT Press.

Pfanzagl, J. (1968). *Theory of measurement.* New York: Wiley.

Pitman, E. J. G. (1938). The estimation of the location and scale parameters of the continuous population of any given form. *Biometrika, 30,* 391.

Ratcliff, R. (1988). A note on mimicking additive reaction time models. *Journal of Mathematical Psychology 32*(2), 192–204.

Roberts, F. S. (1979). *Measurement theory with applications to decision making, utility and social sciences.* Reading, MA: Addison-Wesley.

Roberts, F. S. (1985). Applications of the theory of meaningfulness to psychology. *Journal of Mathematical Psychology, 29,* 311–332.

Roberts, F. S., & Franke, C. H. (1976). On the theory of uniqueness in measurement. *Journal of Mathematical Psychology, 14,* 211–218.

Roberts, F. S., & Marcus-Roberts, H. M. (1987). Meaningless statistics. *Journal of Educational Statistics, 12,* 383–394.

Ross, S. M. (1983). *Stochastic processes.* New York: Wiley.

Rubin, H. (1951). A complete class of decision procedures for distributions with monotone likelihood ratio. *Annals of Mathematical Statistics, 22,* 608.

Schweickert, R. (1978). A critical path generalization of the additive factors method: Analysis of a Stroop task. *Journal of Mathematical Psychology, 18,* 105–139.

Schweickert, R., & Townsend, J. T. (1989). A trichotomy: Interactions of factors prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *Journal of Mathematical Psychology, 33,* 328–347.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology, 32*(2), 135–168.

Sternberg, S. (1973, November). *Evidence against self-terminating memory search from properties of reaction time distributions.* Paper presented at Meeting of Psychonomic Society, St. Louis, MO.

Stevens, S. S. (1951). Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: Wiley.

Stevenson, M. K., Kanfer, F. H., & Higgins, J. M. (1984). Effects of goal specificity and time cues on pain tolerance. *Cognitive Therapy and Research, 8*(4), 415–426.

Suppes, P., & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology,* (Vol. 1). New York: Wiley.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133–186) Hillsdale, NJ: Erlbaum.

Townsend, J. T. (in press). On the proper scales for reaction time. In H. G. Geissler, S. Link, & J. T. Townsend (Eds.), *Cognition, information processing and psychophysics: Basic issues.* Hillsdale, NJ: Erlbaum.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory,* (Vol. 3, pp. 200–239). Hillsdale, NJ: Erlbaum.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge, England: Cambridge University Press.

Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin, 96,* 394–401.

Townsend, J. T., & Roos, R. N. (1973). Search reaction time for single targets in multiletters stimuli with brief visual displays. *Memory & Cognition, 1,* 319–332.

Townsend, J. T., & Schweickert, R. (1985). Interactive effects of factors prolonging processes in latent mental networks. In G. d'Yadewalle (Ed.), *Cognition, information processing and motivation, Proceedings of 23rd International Congress of Psychology, 3,* 255–276. North Holland, Amsterdam: Elsevier.

Townsend, J. T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology, 33,* 309–327.

Van Lier, L. (1989). A simple sufficient condition for the unique representability of a finite qualitative probability by a probability measure. *Journal of Mathematical Psychology, 33,* 91–98.

Vorberg, D. (1981). Reaction time distributions predicted by serial, self-terminating models of memory search. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology. Proceedings of Symposium in Applied Mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics* (pp. 301–318). New York: American Mathematical Society.

Wilks, S. S. (1962). *Mathematical statistics.* New York: Wiley.

Appendix

Theoretical Derivations

Of interest are distributions on the real line. It is well known that every distribution function can be written as a convex combination of (a) an absolutely continuous part, (b) a discrete part, and (c) a continuous singular part (e.g., Chung, 1974, pp. 10–13). Most practical applications involve *a* or *b*, where *a* implies the existence of a continuous density representation and *b* the existence of a discrete probability distribution.

The following proofs are typically proven or referenced on the basis of the existence of continuous density functions. The situation for discrete distributions would be similar. It appears that with appropriate care, analogous propositions and proofs for distributions composed of densities, discrete probabilities, and continuous singular components can be fashioned (Lehmann, 1955), but these seem to have little pragmatic value in psychology.

An additional comment is relevant about the measurement scale, in relation to the operation of integration that is used in some of the proofs: If the scale is ordinal, then the usual definitions of integration do not apply because the limits may not continue to hold after monotonic transformation. There are several ways to confront this problem. One, as suggested by a referee, would be to define a generalized notion of integration that would encompass the new class of functions. That strategy may be difficult and goes beyond the aims of this article. A second, that also may or may not be possible or tractable, would be to refer all candidate functions to a continuous (or absolutely continuous) canonical version, which would then be subjected to the operations of the integral or differential calculus. The third approach is the easiest, but acceptable here, and is in accord with the fact that most sets of real-world data yield discrete distributions. All the theorems have discrete analogues and proofs that thereby apply to actual data and tests that can be used in real-world settings. However, for convenience and readability, I present the continuous versions in this article.

The cases in which a strict order "greater than" rather than the weaker "greater than or equal to", is of most interest. Because of instances in which the measured variate x is bounded (e.g., $F(x) = 1$ for $x \geq x_1$ or $F(x) = 0$ for $x < x_0$) and therefore both $F_A(x)$ and $F_B(x)$ might be equal, say to 1 for a range of x, by convention ">" and "for all x" are used to refer to all values of x except at such end domains where equality may be attained. This strategy does not affect the results. Also, by convention Mean(X) is used to represent the expected value $E(X)$ of a random variable X.

The medians referred to in Propositions 1 and 2 do not necessarily have to be unique in that other examples exist where the means are ordered in one direction and the (possibly nonunique) medians in another, or that $P(X_A \leq X_B) \leq \frac{1}{2}$. However, the particular examples in these propositions do obey uniqueness of medians, which is implied by the densities being everywhere positive, or other conditions also suffice.

Proposition 1

$[M(X_A) < M(X_B)]$ neither implies nor is implied by $[Mdn(X_A) < Mdn(X_B)]$.

Proof

To prove that $M(X_A) < M(X_B)$ does not imply $Mdn(X_A) < Mdn(X_B)$, let $f_A(x)$ be normal with a mean of 4 and variance of 0.5 and $f_B(x)$ be exponential with rate parameter 0.2. It will be seen that $Mdn(X_B) = 3.47 < Mdn(X_A) = 4$, whereas $M(X_B) = 5 > M(X_A) = 4$. Now, interchange X_A, X_B.

Proposition 2

A. $[M(X_A) < M(X_B)]$ neither implies nor is implied by $[P(X_A \leq X_B) > \frac{1}{2}]$.

Proof

That $P(X_A \leq X_B) > \frac{1}{2}$ does not imply $M(X_A) < M(X_B)$ was proven in Townsend and Ashby (1978, pp. 227–228). This was accomplished by way of an example in which $P(X_A \leq X_B) > \frac{1}{2}$, but $M(X_A) > M(X_B)$. Now, interchange X_A, X_B.

B. $[P(X_A \leq X_B) > \frac{1}{2}]$ neither implies nor is implied by $[Mdn(X_A) < Mdn(X_B)]$.

This is shown for the discrete case in Townsend and Ashby (1978, p. 228). This counterexample is easily converted to a proof involving densities by using each discrete probability mass as a weight for a density with sufficiently small variance. This yields two probability mixture densities having the desired properties, namely, $P(X_A \leq X_B) > \frac{1}{2}$ but $Mdn(X_A) > Mdn(X_B)$.

The first part of the next proposition is proven under the assumption that the distribution functions are strictly increasing. However, a proof suggested by Hans Colonius is given in Townsend and Schweickert (1989) that avoids that assumption.

Proposition 3

A. If the distribution functions are strictly increasing and $F_A(x) > F_B(x)$ for all x, then there exists a positive random variable X_C, possibly dependent on X_A, such that the new random variable $X_A + X_C$ has the same distribution function as X_B.

B. Adding a positive random variable to X_A creates a new random variable X_B with the property that $F_A(x) > F_B(x)$ for all x.

Proof of A and B

Both A and B are proven by Townsend and Schweickert (1989) for positive-valued random variables X_A, X_B. However, as was suggested by a referee, it seems a good idea to include the proofs here for the more general case. As is noted in the text, it helps to avoid problems to define another random variable X^*_A with the property that its marginal distribution, integrating over X_C, is identical to that of X_A. Thus, if $f_{A,C}(x^*_A, x_C)$ is the joint density on X^*_A, X_C, then $f_{A^*}(x) = \int_{x_C=-\infty}^{\infty} f_{A^*,C}(x_A, x_C)\, dx_C = f_A(x)$ for all x contained in $(-\infty, +\infty)$.

This device precludes incorrect conclusions such as $P(X_A < X_B) = P(X_A < X_A + X_C) = P(0 < X_C) = 1$. Indeed, X_A and X_B are usually associated with independent samples in which the foregoing consequence would be untenable. Because X_A and X^*_A are typically independent, $P(X_A < X^*_A + X_C)$ is not necessarily equal to 1. Understanding this, it should do no harm to simplify my notation and use X_A in place of X^*_A.

Proof of A

Suppose $F_A(x) > F_B(x)$ for all x contained in $(-\infty, +\infty)$. If Y is a random variable with distribution function $F(y)$, then $F(Y)$ is a random variable with the uniform distribution; that is, $P[F(Y) \le y] = y$ where $0 \le y \le 1$. Define $X_C = F_B^{-1}[F_A(X_A)] - X_A$. Then X_C is positive with probability 1 because $F_A(x) > F_B(x)$. Furthermore, $P(X_A + X_C \le x) = P[F_B^{-1}[F_A(X_A)] \le x] = P[F_A(X_A) \le F_B(x)] = F_B(x)$, as desired.

Proof of B

Suppose $X_B = X_A + X_C$, and $P(X_C = 0) = 0$. Then

$$P(X_A + X_C \le x) = P(X_A \le x - X_C)$$

$$= \int_{-\infty}^{\infty} P(X_A \le x - x_C | X_C = x_c) f_{X_C}(x_C)\, dx_C$$

$$< \int_{-\infty}^{\infty} P(X_A \le x | X_C = x_C) f_{X_C}(x_C)\, dx_C = P(X_A \le x),$$

under the hypothesis. Hence, $F_B(x) = P(X_A + X_C \le x) < P(X_A < x) = F_A(x)$.

Observe (relevant to A) that X_C is perfectly correlated with X_A in that if $X_A = x$, then a fortiori $X_C = F_B^{-1}[F_A(x)] - x$. This rules out X being negative, for instance. However, the conditional distribution on X_C, given $X_A = x$, is discontinuous because of its being located at $F_B^{-1}[F_A(x)] - x$ with probability 1. Notice also that there is no guarantee of uniqueness here. Thus, this method imposed on a convolution of densities of two independent random variables will not, in general, return the original random variables. As a quick example of 3A, suppose $F_A(x) = 1 - e^{-Ax}$ for $x \ge 0$ and is zero otherwise, and let $F_B(x) = 1 - e^{-Bx}$ for the same domain with $A > B > 0$. Then X_C is (marginally) exponentially distributed with

$$F_C(x) = 1 - \exp\left[\frac{(-AB)t}{A - B}\right].$$

Proposition 4

A. $F_A(x) > F_B(x)$ for all x implies all, but is not implied by any, of the following: (i) $M(X_A) < M(X_B)$. (ii) $Mdn(X_A) < Mdn(X_B)$, if F_A, F_B are strictly increasing. (iii) If X_A and X_B are also independent, then $P(X_A \le X_B) > \frac{1}{2}$.

Proof

(i) Let $\bar{F}_i(x) = 1 - F_i(x)$ = survivor function for X_i. Then

$$M(X_A) = \int_0^{\infty} \bar{F}_A(x)\, dx - \int_{-\infty}^{0} F_A(x)\, dx$$

versus

$$M(X_B) = \int_0^{\infty} \bar{F}_B(x)\, dx - \int_{-\infty}^{0} F_B(x)\, dx$$

leads to

$$M(X_B) - M(X_A) = \int_{-\infty}^{\infty} [F_A(x) - F_B(x)]\, dx > 0,$$

as was to be shown.

(ii) Because $F_A(x) > F_B(x)$ for all x, and both being strictly increasing,

$$Mdn(X_A) = F_A^{-1}(\tfrac{1}{2}) < F_B^{-1}(\tfrac{1}{2}) = Mdn(X_B).$$

If the medians are not unique, then it could happen that $\text{lub}[x: F_A(x) = \frac{1}{2}]$ and $\text{glb}[x: F_B(x) = \frac{1}{2}]$ are unequal. However, it still could not occur that an $x = x_0$ existed such that $F_A(x_0) = \frac{1}{2}$ and $F_B(x_0) = \frac{1}{2}$, because then obviously $F_A(x) \not> F_B(x)$ for all x.

(iii) This is proven in Townsend and Ashby (1978, p. 227) for positive valued random variables, and the same type of proof works as well on the entire real line.

B. If X_A and X_B are dependent, then $P(X_A \le X_B) > \frac{1}{2}$ is not implied by the ordering $F_A(x) > F_B(x)$.

Proof

For simplicity, the example is discrete, but the structure would be similar for a continuous counterpart. Let

$$f_{AB}(x_A, x_B) = 0.2 \text{ if } x_A = 1, x_B = 1.5;$$

$$f_{AB}(x_A, x_B) = 0.2 \text{ if } x_A = 2, x_B = 2.5;$$

$$f_{AB}(x_A, x_B) = 0.2 \text{ if } x_A = 3, x_B = 3.5;$$

$$f_{AB}(x_A, x_B) = 0.4 \text{ if } x_A = 4, x_B = 0.5;$$

and

$$f_{AB}(x_A, x_B) = 0 \text{ elsewhere}$$

when f_{AB} is the joint probability mass function.

Proposition 5

Suppose there are at most a finite number of density crossovers. Then, $F_A(x) > F_B(x)$ implies that either (a) there are no density crossovers, that is, there is no point $x = x_0$ such that $f_A(x_0) = f_B(x_0) > 0$, $f_A(x) > f_B(x)$ if $x < x_0$, and $f_A(x) < f_B(x)$ if $x > x_0$; or (b) there exists $x_1, x_2, \ldots, x_n$ with n odd such that $f_A(x_i) = f_B(x_i) > 0$ and otherwise $f_A(x) \ne f_B(x)$.

Proof

This proposition is most easily proven through the contrapositive, that is, an even (nonzero) number of crossover points implies that there exists at least one x such that $F_A(x) = F_B(x)$; that is, the distribution functions cannot be ordered. Suppose without loss of generality that $n \geq 2$ is even and that for all x contained in an interval $(-\infty, x_1)$, $f_A(x) > f_B(x)$. Then because n is even, on the interval $(x_n, +\infty)$, $f_A(x) > f_B(x)$ again. It follows that for x contained in $(-\infty, x_1)$, $F_A(x) > F_B(x)$. However, for x contained in $(x_n, +\infty)$, the sum total area left in the tail of the X_A density must be greater than that for the X_B density, which implies that on this interval, $F_A(x) < F_B(x)$. Therefore the Fs are not ordered in contradiction to the hypothesis.

Proposition 6

Again, suppose there are at most a finite number of density crossovers.

A. If there are no density crossovers and f_A, f_B are not equivalent, then the distribution functions are ordered.

Proof

Suppose without loss of generality that there exists an $x = x_0$ such that $f_A(x_0) > f_B(x_0)$ and $f_A(x) \geq f_B(x)$ for all $x < x_0$. Obviously, $F_A(x_0) > F_B(x_0)$, and because of continuity and the fact that there are no density crossovers, this ordering remains for all x.

B. If the densities cross exactly once, then $F_A(x) > F_B(x)$ for all x.

Proof

This is proven in Townsend and Ashby (1978, pp. 282–283). It follows intuitively because early on, if $f_A(x) > f_B(x)$, then also $F_A(x) > F_B(x)$ for small x. But, because of a single crossover point, F_A and F_B never have a chance to reverse themselves.

C. If the densities f_A, f_B cross an odd number of times but greater than 1, then the respective distributions may or may not be ordered.

Proof

An example of two densities that cross an odd number of times yet do not force $F_A(x) > F_B(x)$ for all x would be tedious to construct in detail, but one can readily set up the mathematics to show how such examples can be formulated. I use three density crossovers for simplicity, but the argument is general.

Because of the three density crossovers at x_1, x_2, x_3, one can effectively segregate the contributions to $F_A(x) - F_B(x)$ into the following intervals: $f_A(x) > f_B(x)$ for $-\infty < x < x_1$; $f_A(x) < f_B(x)$ for $x_1 < x < x_2$; $f_A(x) > f_B(x)$ for $x_2 < x < x_3$; $f_A(x) < f_B(x)$ for $x_3 < x < +\infty$, and $f_A(x) = f_B(x)$ for x_1, x_2, x_3. It is evident that $F_A(x) > F_B(x)$ for $-\infty < x \leq x_1$, and it is evident that one can produce $F_A(x) < F_B(x)$ when $x_1 < x < x_2$ by simply arranging

$$\int_{-\infty}^{x_1} [f_A(x) - f_B(x)]\, dx$$

and

$$\int_{x_1}^{x_2} [f_B(x) - f_A(x)]\, dx$$

so that the latter is sufficiently large in relation to the former to make $F_B(x) - F_A(x) > 0$ for some x contained in (x_1, x_2). This completes the proof.

On the other hand, the order $F_A(x) > F_B(x)$ will be preserved everywhere if

$$\int_{x_1}^{x_2} [f_B(x) - f_A(x)]\, dx$$

is kept small in relation to

$$\int_{-\infty}^{x_1} [f_A(x) - f_B(x)]\, dx.$$

Furthermore, if $F_A(x_2) > F_B(x_2)$ still, then $F_A(x) > F_B(x)$ for all x, because certainly $F_A(x_3) > F_B(x_3)$ (because $f_A(x) > f_B(x)$ for $x_2 < x < x_3$) and then $1 - F_A(x) < 1 - F_B(x)$ for $x_3 < x < +\infty$ (because the tail area in f_A is smaller than that for f_B) so that $F_A(x) > F_B(x)$ there also. In the case of odd $n > 3$ to preserve the order $F_A(x) > F_B(x)$ for all x, it is of course necessary to ensure that the intervals where $f_B(x) > f_A(x)$ do not produce a sum of integrated negative values at any x, that is greater than the sum of integrated positive values of $f_A(x) - f_B(x)$; otherwise, the order is broken.

D. If the densities cross an even number of times greater than zero, the distribution functions cannot be ordered.

Proof

This statement is logically equivalent to Proposition 5.

Proposition 7

The distribution function for $X_B = X_A + c$, where c is a positive constant, implies $F_A(x) > F_B(x)$, but the converse does not hold.

Proof

Note that $F_B(x) = P(X_A + c \leq x) = P(X_A \leq x - c) < P(X_A \leq x) = F_A(x)$. However, any nonshifted but ordered pair of distributions obeys $F_A(x) > F_B(x)$ but not $X_B = X_A + c$. For instance, $F_A(x) = 1 - e^{-Ax}$, $F_B(x) = 1 - e^{-Bx}$ for $x \geq 0$, and $A > B > 0$ is an example.

Proposition 8

If the hazard functions are ordered, then so are the distribution functions, but not vice versa.

Proof

For distributions on $x \geq 0$, the fact that $[H_A(x) > H_B(x)]$ implies $[F_A(x) > F_B(x)]$ is demonstrated by Townsend and

Ashby (1978, pp. 224–225), as is the fact that $F_A(x) > F_B(x)$ does not imply $[H_A(x) > H_B(x)]$. Again, the arguments go through without difficulty when x is contained in $(-\infty, +\infty)$.

Proposition 9

A shift pair of distributions, $f_B(x) = f_A(x - c)$ for $c > 0$ (a constant), does not force the hazard functions to be ordered, or vice versa.

Proof

If $H(x)$ is monotonically decreasing in some interval, then $H(x) < H(x - c)$, so the ordering is the reverse of what it should be. This proves the first part. The second part follows from this, noting that for, say,

$$f_A(x) = \frac{A(Ax)^{k-1}e^{-Ax}}{(k-1)!}$$

and

$$f_B(x) = \frac{B(Bx)^{k-1}e^{-Bx}}{(k-1)!}$$

for $x \geq 0$, and $A > B > 0$ fixed parameters, $H_A(x) > H_B(x)$ for all $x > 0$, yet f_B is not a shift version of f_A.

Proposition 10

A pair of ordered hazard functions $H_A(x) > H_B(x)$ does not imply that their densities cross exactly once.

Proof

A counterexample is given with $H_A(x) > H_B(x)$ and densities that possess three crossovers. Thus, the distribution functions are ordered, but $L(x)$ is not monotonically decreasing. Let $f_A(x)$ and $f_B(x)$ be defined on $[0, +\infty)$ and define $H_A(x)$ and $H_B(x)$ so that $H_A(0) > H_B(0)$, and thus $L(0) > 1$, followed by $L(x)$ decreasing to $L(x_1) < 1$ at $x = x_1$. This implies a crossover at some $x_0 < x_1$. Next, let $H_A(x)$ and $H_B(x)$ go to values less than 1 with a lower bound on $H_A(x)$ and with $H_B(x)$ becoming very small. With properly chosen values, the likelihood ratio

$$L(x) = \left(\frac{H_A(x)}{H_B(x)}\right) \exp\left\{-\int_0^x [H_A(x') - H_B(x')]\,dx'\right\}$$

retreats back to $L(x) > 1$. This is because $H_A(x)/H_B(x)$ becomes very large whereas the exponential term is bounded above a sufficiently large number that the product grows larger than 1. Finally, pick functions $H_A(x)$, $H_B(x)$ to take off from some x such that $L(x) > 1$ and such that $L(x)$ decreases in a monotonic fashion, thereby producing the final crossover point. All this may be done with continuous and differentiable H_A and H_B hazard functions. I am content with continuity in this example, which I leave the reader to work out in greater detail.

Basically, $x_0 < x_1 < x_2 < x_3 < x_4$ where x_0 = first crossover point, x_1 = the first point where $L(x)$ begins to climb, x_2 = second crossover point, x_3 = point where $H_B(x)$ begins to increase again, and x_4 = point in example where $H_B(x)$ attains its final value for the rest of $x \geq x_4$. A and B are positive constants with $A > B$. The values of x_0 and x_1 will be related to A, B and how far below 1 one wishes $L(x)$ to go before reversal. Similar comments can be made with regard to the other designated points of the variate. Define $f_A(x)$ and $f_B(x)$ through their hazard functions:

$$H_A(x) = 0, \quad x < 0,$$

$$H_A(x) = A, \quad 0 \leq x,$$

$$H_B(x) = 0, \quad x < 0,$$

$$H_B(x) = B, \quad 0 \leq x \leq x_1,$$

$$H_B(x) = \left(\frac{x_3 - x}{x_3 - x_1}\right)B + \left(\frac{x - x_1}{x_3 - x_1}\right)\frac{Ae^{-(x_3-x_1)}e^{-(A-B)x_1}}{1+\epsilon},$$

$$x_1 < x \leq x_3$$

$$H_B(x) = \left(\frac{x_4 - x}{x_4 - x_3}\right)\frac{Ae^{-(x_3-x_1)}e^{-(A-B)x_1}}{1+\epsilon} + \left(\frac{x - x_3}{x_4 - x_3}\right)B,$$

$$x_3 < x \leq x_4$$

$$H_B(x) = B, \quad x_4 < x.$$

The only parameter as yet unexplained is ϵ, which equals the amount above 1 that $L(x)$ goes after crossing 1 the second time. A set of numbers that works in the preceding example is $A = 1$, $B = \frac{1}{2}$, $x_1 = 2.77$, $x_3 = 4.77$, and $\epsilon = 1$. The x_4 is arbitrary as long as it is greater than x_3, and x_0 and x_2 may be calculated from the preceding information.

Proposition 11

A strictly decreasing likelihood function $L(x) = f_A(x)/f_B(x)$ (defined for all $f_B(x) \neq 0$) implies that $H_A(x) > H_B(x)$, but not vice versa.

Proof

The first part is shown in Townsend and Ashby (1983, pp. 281–282), with an obvious extension to $t \in (-\infty, +\infty)$. The fact that the implication does not reverse can be concluded with the following example: Let $f_A(x) = 2xe^{-x^2}$, $H_A(x) = 2x$ for $x \geq 0$ and 0 otherwise, and let $f_B(x) = xe^{-x}$, $H_B(x) = x/(1 + x)$ for $x \geq 0$ and 0 otherwise. It can be seen that $H_A(x)/H_B(x) = 2(x+1) > 1$, so $H_A(x) > H_B(x)$ for all (defined) x. However, $L(x) = f_A(x)/f_B(x) = 2e^{-x^2+x}$ is clearly a nonmonotonic function of x.

Proposition 12

A strictly decreasing $L(x)$ implies a single crossover point for $f_A(x)$ and $f_B(x)$ in "reasonable" cases.

Proof

Those cases seem to be when $f_A(x)$ and $f_B(x)$ can both be nonzero on an interval of their definition, yet $f_A(x) > f_B(x)$ for sufficiently small x. Then apparently the decreasing $L(x)$ forces a unique crossover, say at x_0 as long as nothing strange happens to f_A or f_B (e.g., $f_B = 0$) before x_0 is reached.

Proposition 13

A strictly decreasing monotone likelihood ratio does not imply that $f_B(x) = f_A(x - c)$ for $c > 0$, a constant.

Proof

An example is found with

$$f_A(x) = \frac{(Ax)^{k-1} A e^{-Ax}}{(k-1)!},$$

where k = positive integer and $A > 0$, for $0 \le x$;

$$f_A(x) = 0, \qquad \text{otherwise;}$$

$$f_B(x) = \frac{(Bx)^{k-1} B e^{-Bx}}{(k-1)!},$$

where k = positive integer and $A > B > 0$ for $0 \le x$;

$$f_B(x) = 0, \qquad \text{otherwise.}$$

$L(x)$ decreases strictly monotonically, but $f_B(x)$ is not a shift of $f_A(x)$.

Proposition 14

The shift family of normals with equal variance produces $L(x)$ strictly decreasing for any two members $f_A(x)$, $f_B(x) = f_A(x - c)$ with $c > 0$.

Proof

This follows immediately from examination of (letting $\sigma = 1$ without loss of generality)

$$\ln L(x) = \ln\left(\frac{f_A(x)}{f_B(x)}\right)$$

$$= \frac{-(x-\mu)^2}{2} + \frac{(x-\mu-c)^2}{2} = \frac{c^2 - 2c\mu - 2cx}{2},$$

a strictly decreasing function of x.

Received August 19, 1988
Revision received March 26, 1990
Accepted April 2, 1990 ■