The Annals of Statistics
1979, Vol. 7, No. 2, 349–371

COMPOSITION RULES FOR PROBABILITIES FROM PAIRED COMPARISONS[1]

BY ROBERT B. LATTA

University of Kentucky

Let p_{ab} be the probability that the outcome of a paired comparison involving a and b is favorable to a. This paper discusses composition rules that generate p_{ac} given p_{ab} and p_{bc}. The basic properties of composition rules are developed via an axiomatic approach.

If $p_{ab} = F(w_a - w_b)$, where F is a distribution function for a distribution that is symmetric about zero, then the paired comparison model is a linear model that is based on the distribution function F. It is shown that given any composition rule, which obeys certain basic axioms, there exists a linear model that generates an identical composition rule.

The behavior of the composition rules are used to place a partial ordering on the paired comparison models and in particular on the linear models. This partial ordering is denoted as the extreme partial ordering. It is shown that linear models based on distributions with short tails tend to be more extreme than those based on distributions with long tails. The resulting partial ordering includes the result that the Thurstone-Mosteller model is more extreme than the Bradley-Terry model.

The extreme ordering can also be used to place a partial ordering on distributions according to the lengths of their tails. The relation of this ordering to the s-ordering and r-ordering is examined.

1. Introduction and summary. Let p_{ab} be the probability that subject a is chosen over subject b (team a beats team b, x_a is larger than x_b, etc.). The knowledge of p_{ab} and p_{bc} should convey some information concerning p_{ac}. This paper is primarily concerned with composition rules that generate p_{ac} given p_{ab} and p_{bc}.

The procedure of comparing two items at a time is known as the method of paired comparisons. In a discussion of a paper by Bradley (1976), David presents a hierarchy of paired comparison models and notes what they imply about the composition rules.

A linear model for paired comparisons (c.f. David [1963] page 13) holds when $p_{ab} = F(w_a - w_b)$, where F can be any distribution function which has the property that $F(x) = 1 - F(-x)$. If we assume that $F^{-1}(p)$ exists and is unique for $p \in (0, 1)$, then $p_{ac} = F[F^{-1}(p_{ab}) + F^{-1}(p_{bc})]$ is the resulting composition rule. There are several special cases of composition rules determined by linear models that are of interest.

A composition rule can be viewed as a function that maps $(0, 1) \times (0, 1)$ into $[0, 1]$. An axiomatic formulation of composition rules will be given and basic

Received June 1976; revised October 1977.

[1]This paper is based on part of the author's doctoral dissertation at Stanford University. Research supported by NIH Grant No. 5 TI GM25-18.

AMS 1970 *subject classification.* Primary 62J15.

Key words and phrases. Paired comparisons, Bradley-Terry model, Thurstone-Mosteller model, s-ordering, r-ordering.

properties of composition rules will be developed. An argument due to Debreu (1958) is developed to show that under general assumptions all composition rules can be derived from linear models.

The behavior of composition rules is used to place a partial ordering on the linear models. The linear model based on F_1 is defined to be more extreme than the linear model based on F_2 if, whenever p_{ab} and p_{bc} are elements of $(\frac{1}{2}, 1)$, then the p_{ac} generated by the F_1 model is greater than or equal to the p_{ac} generated by the F_2 model. It is shown that the model based on the uniform distribution is more extreme than any linear model that is based on a distribution function which has a unimodal density. Further, models based on distribution functions with short tails tend to be more extreme than models based on distribution functions with long tails. These results are applied to several linear models, inducing a partial ordering on them. This partial ordering includes the result that the Thurstone-Mosteller model is more extreme than the Bradley-Terry model.

The above partial ordering for linear models corresponds to a partial ordering for distribution functions according to the length of their tails. It is shown that the r-ordering of Lawrence (1975) implies this ordering. In other words, if F_1 is r-ordered with respect to F_2 then the linear model based on F_1 is more extreme than the linear model based on F_2.

2. Axiomatic formulation of composition rules. In this section composition rules are formally defined and several properties of composition rules are presented. Let S be the abstract set representing the collection of possible subjects.

DEFINITION. p_{ab} is the probability that the outcome of a paired comparison involving $a, b \in S$ is favorable to a.

AXIOM 2.1. (No ties). $p_{ab} + p_{ba} = 1$ for every $a, b \in S$.

AXIOM 2.2. (Richness of S). For every $\alpha \in (0, 1)$ and $a \in S$ there exists $b \in S$ such that $p_{ab} = \alpha$.

Axiom 2.2 implies that S cannot be countable. Usually, in applications, only a finite subset of S and probability relationships that can be generated in a simple way are of interest.

DEFINITION. A composition rule is a function G that maps $(0, 1) \times (0, 1)$ into $[0, 1]$ with the property that $p_{ac} = G(p_{ab}, p_{bc})$ for every $a, b, c \in S$ such that $p_{ab}, p_{bc} \in (0, 1)$.

PROPERTY 2.1. $G(p_1, p_2) = 1 - G(1 - p_2, 1 - p_1)$ *for* $p_1, p_2 \in (0, 1)$.

PROOF. By Axiom 2.2 there exists $a, b, c \in S$ such that $p_1 = p_{ab}$ and $p_2 = p_{bc}$. By Axiom 2.1 $p_{ba} = 1 - p_1$ and $p_{cb} = 1 - p_2$.

$$p_{ac} = G(p_{ab}, p_{bc}) = G(p_1, p_2)$$

$$p_{ca} = G(p_{cb}, p_{ba}) = G(1 - p_2, 1 - p_1).$$

Hence

$$G(p_1, p_2) = p_{ac} = 1 - p_{ca} = 1 - G(1 - p_2, 1 - p_1). \qquad \square$$

PROPERTY 2.2. *$G(p_1, 1 - p_1) = \frac{1}{2}$ for $p_1 \in (0, 1)$.*

PROOF. By Property 2.1, $G(p_1, 1 - p_1) = 1 - G(p_1, 1 - p_1)$. $\square$

PROPERTY 2.3. *For $p_1, p_2, p_3 \in (0, 1)$*

$$p_1 = G(p_3, 1 - p_2) \Leftrightarrow p_3 = G(p_1, p_2) \Leftrightarrow p_2 = G(1 - p_1, p_3).$$

PROOF. First assume $p_3 = G(p_1, p_2)$. There exists $a, b, c \in S$ such that $p_{ab} = p_1$ and $p_{bc} = p_2$. Then

$$p_{ac} = G(p_1, p_2) = p_3.$$

Also

$$p_2 = p_{bc} = G(p_{ba}, p_{ac}) = G(1 - p_1, p_3)$$
$$p_1 = p_{ab} = G(p_{ac}, p_{cb}) = G(p_3, 1 - p_2).$$

Now assume $p_2 = G(1 - p_1, p_3)$. Then there exists $a, b, c \in S$ such that $p_{ba} = 1 - p_1$ and $p_{ac} = p_3$.

Hence $p_2 = G(1 - p_1, p_3) = G(p_{ba}, p_{ac}) = p_{bc}$.
Therefore $p_3 = p_{ac} = G(p_{ab}, p_{bc}) = G(p_1, p_2)$.
Similarly $p_1 = G(p_3, 1 - p_2) \Rightarrow p_3 = G(p_1, p_2)$. $\square$

PROPERTY 2.4. *For $p \in (0, 1)$, $G(\frac{1}{2}, p) = p$ and $G(p, \frac{1}{2}) = p$.*

PROOF. The result follows from Properties 2.2 and 2.3. $\square$

It now becomes obvious that the existence of a composition rule imposes a restriction on the interdependence of the values of p_{ab}. The following conditions also appear to be reasonable:

AXIOM 2.3. (Symmetry). $G(p_1, p_2) = G(p_2, p_1)$.

AXIOM 2.4. (Monotonicity). Assume $p_1, p_2, p_3 \in (0, 1)$. Then $p_1 > p_3 \Rightarrow G(p_1, p_2) \geqslant G(p_3, p_2)$. If, in addition, $G(p_1, p_2) \in (0, 1)$, then $p_1 > p_3 \Leftrightarrow G(p_1, p_2) > G(p_3, p_2)$.

PROPERTY 2.5. (*Continuity*). *$G(p_1, p_2)$ is continuous on $(0, 1) \times (0, 1)$.*

PROOF. Let $\alpha < \beta$ and define

$$G^{-1}[(\alpha, \beta)] = \{(p_1, p_2) \in (0, 1) \times (0, 1) | G(p_1, p_2) \in (\alpha, \beta)\}.$$

It is sufficient to show that $G^{-1}[(\alpha, \beta)]$ is open. If $(\alpha, \beta) \cap [0, 1] = \varnothing$, then $G^{-1}[(\alpha, \beta)] = \varnothing$ and hence is open. Now assume $(\alpha, \beta) \cap [0, 1] \neq \varnothing$; then $(\alpha, \beta) \cap (0, 1) \neq \varnothing$. Let $p \in [(\alpha, \beta) \cap (0, 1)]$. By Property 2.4 $G(p, \frac{1}{2}) = p$. Hence $G^{-1}[(\alpha, \beta)] \neq \varnothing$. Let $(p_1', p_2') \in G^{-1}[(\alpha, \beta)]$. Assume $G(p_1', p_2') = \gamma$. Then $\alpha < \gamma < \beta$.

It is now sufficient in this case to show that (p_1', p_2') is an interior point of $G^{-1}[(\alpha, \beta)]$.

First assume $\gamma \in (0, 1)$. Choose γ_1 and γ_2 such that $\max\{0, \alpha\} < \gamma_1 < \gamma_2 < \gamma$. By the richness Axiom 2.2 there exists a, b, c, d, e elements of S such that $p_{ab} = p_1'$, $p_{bc} = p_2'$, $p_{ad} = \gamma_2$, $p_{ed} = \gamma_1$. Therefore $G(p_{ab}, p_{bd}) = p_{ad} = \gamma_2$ and $G(p_{eb}, p_{bd}) = p_{ed} = \gamma_1$. Axiom 2.4 and $\gamma_1 < \gamma_2 < \gamma$ imply that $p_{eb} < p_{ab}$ and $p_{bd} < p_{bc}$. Let $p_1^* = p_{eb}$ and $p_2^* = p_{bd}$, then $p_1 > p_1^*$ and $p_2 > p_2^*$ imply $G(p_1, p_2) > \gamma_1$.

Similarly let $\gamma < \gamma_3 < \min\{1, \beta\}$. Then there exists $p_1^{**} > p_1', p_2^{**} > p_2'$ such that $p_1 < p_1^{**}$ and $p_2 < p_2^{**}$ imply $G(p_1, p_2) < \gamma_3$. Hence $(p_1, p_2) \in (p_1^*, p_1^{**}) \times (p_2^*, p_2^{**})$ implies $G(p_1, p_2) \in (\gamma_1, \gamma_3) \subset (\alpha, \beta)$. Therefore (p_1', p_2') is an interior point of $G^{-1}[(\alpha, \beta)]$.

Now assume $\gamma = 0$. Then, by the above argument if $0 < \gamma_3 < \min\{1, \beta\}$, there exists p_1^{**} and p_2^{**} such that $p_1 < p_1^{**}$ and $p_2 < p_2^{**}$ imply $G(p_1, p_2) < \gamma_3$. Axiom 2.4 implies

$$G(p_1'/2, p_2'/2) = 0.$$

Hence $p_1'/2 < p_1 < p_1^{**}$ and $p_2'/2 < p_2 < p_2^{**}$ imply

$$G(p_1, p_2) \in [0, \gamma_3) \subset (\alpha, \beta).$$

Hence (p_1', p_2') is an interior point of $G^{-1}[(\alpha, \beta)]$. The case where $\gamma = 1$ is handled similarly. □

PROPERTY 2.6. (a) $p_1, p_2 \in [\frac{1}{2}, 1) \Rightarrow G(p_1, p_2) \geqslant \max\{p_1, p_2\}$;
(b) $p_1, p_2 \in (0, \frac{1}{2}] \Rightarrow G(p_1, p_2) \leqslant \min\{p_1, p_2\}$.

PROOF. (a) By Axiom 2.4 and Property 2.4,

$$G(p_1, p_2) \geqslant G(\tfrac{1}{2}, p_2) = p_2$$
$$G(p_1, p_2) \geqslant G(p_1, \tfrac{1}{2}) = p_1.$$

(b) is proved similarly. □

By virtue of Property 2.6, a composition rule exists only if the paired comparison model is strongly transitive (Brunk [1960]). Nevertheless, models which are weakly transitive have received considerable attention (e.g. de Cani [1969, 1972], Remage and Thompson [1966], and Thompson and Remage [1964]) and are important in certain applications (c.f. David [1963] page 13).

PROPERTY 2.7. (a) $1 > p_1 \geqslant 1 - p_2 \geqslant \frac{1}{2} \Rightarrow \frac{1}{2} \leqslant G(p_1, p_2) \leqslant p_1$;
(b) $0 < p_1 \leqslant 1 - p_2 \leqslant \frac{1}{2} \Rightarrow \frac{1}{2} \geqslant G(p_1, p_2) \geqslant p_1$.

PROOF. The result follows from Property 2.4 and Property 2.6. □

DEFINITION. Two composition rules are identical if $G_1(p_1, p_2) \equiv G_2(p_1, p_2)$ for all $(p_1, p_2) \in [(0, 1) \times (0, 1)]$.

PROPERTY 2.8. *If* $G_1(p_1, p_2) = G_2(p_1, p_2)$ *in one of the quadrants* $(0, \frac{1}{2}) \times (0, \frac{1}{2})$, $(0, \frac{1}{2}) \times (\frac{1}{2}, 1)$, $(\frac{1}{2}, 1) \times (0, \frac{1}{2})$, *or* $(\frac{1}{2}, 1) \times (\frac{1}{2}, 1)$ *then* G_1 *and* G_2 *are identical.*

PROOF. Assume $G_1(p_1, p_2) = G_2(p_1, p_2)$ on $(\frac{1}{2}, 1) \times (\frac{1}{2}, 1)$. If $p_3 = \frac{1}{2}$ or $p_4 = \frac{1}{2}$, then by Property 2.4, $G_1(p_3, p_4) = G_2(p_3, p_4)$. If $(p_3, p_4) \in [(0, \frac{1}{2}) \times (0, \frac{1}{2})]$ then

$$\begin{aligned} G_1(p_3, p_4) &= 1 - G_1(1 - p_4, 1 - p_3) && \text{by Property 2.1;} \\ &= 1 - G_2(1 - p_4, 1 - p_3) && \text{by assumption;} \\ &= G_2(p_3, p_4) && \text{by Property 2.1.} \end{aligned}$$

If $\frac{1}{2} < 1 - p_3 \leqslant p_4 < 1$ then $G_1(p_3, p_4) \in [\frac{1}{2}, p_4)$ by Property 2.7. Let $G_1(p_3, p_4) = p_5$ then

$$\begin{aligned} p_4 &= G_1(1 - p_3, p_5) && \text{by Property 2.3;} \\ &= G_2(1 - p_3, p_5) && \text{by assumption.} \end{aligned}$$

Therefore

$$p_5 = G_2(p_3, p_4) \qquad \text{by Property 2.3.}$$

If $\frac{1}{2} < p_4 \leqslant 1 - p_3 < 1$ then

$$\begin{aligned} G_1(p_3, p_4) &= 1 - G_1(1 - p_4, 1 - p_3) && \text{by Property 2.1;} \\ &= 1 - G_2(1 - p_4, 1 - p_3) && \text{by the case above;} \\ &= G_2(p_3, p_4) && \text{by Property 2.1.} \end{aligned}$$

If $p_4 \leqslant \frac{1}{2} \leqslant p_3$ then using Property 2.1 and the results above will show that $G_1(p_3, p_4) = G_2(p_3, p_4)$. Hence G_1 and G_2 are identical. The case of assuming equality in one of the other quadrants is handled similarly. □

Figure 2.1 (Debreu, 1958) can be used to illustrate the meaning of some of the above axioms and properties. The dotted lines are isoprobability or contour lines of $G(p_1, p_2)$. In other words I_α is the set of all points (p_1, p_2) such that $G(p_1, p_2) = \alpha$.

(a) Axiom 2.4 implies that when $\alpha \in (0, 1)$, then I_α intersects the line $p_i = p$, $i = \{1, 2\}$ at most one time.

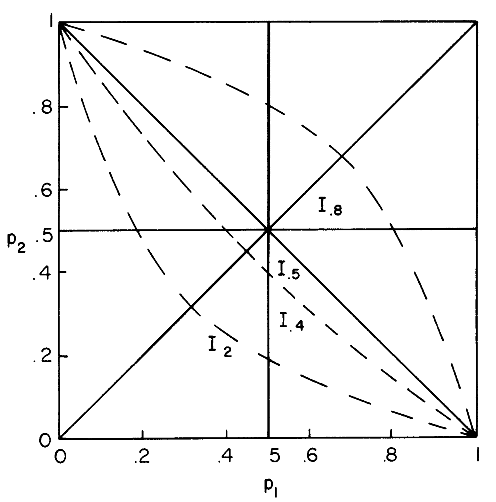


FIG. 2.1. Isoprobability lines for $G(p_1, p_2)$.

(b) Axiom 2.4 and Property 2.5 together imply that I_α is a continuous line for $\alpha \in (0, 1)$.
(c) Property 2.1 implies that I_α is the reflection of $I_{1-\alpha}$ through $(\frac{1}{2}, \frac{1}{2})$.
(d) Property 2.2 implies that $I_{.5}$ is the straight line from (0, 1) to (1, 0).
(e) Property 2.4 implies that I_α passes through $(\alpha, \frac{1}{2})$ and $(\frac{1}{2}, \alpha)$.
(f) Axiom 2.3 implies that I_α is the reflection of itself through the diagonal line $p_1 = p_2$.
(g) Axiom 2.4 implies that I_α and I_β do not intersect if $\alpha \neq \beta$.
(h) Property 2.7 implies that when $\alpha \in (0, 1)$, then I_α approaches (1, 0) and (0, 1).

3. Composition rules for linear models.

DEFINITION. A linear model for paired comparisons (c.f. David [1963] page 13) holds if, for each $s \in S$, there exists a real valued constant W_s, if there exists a distribution function F such that $F(x) = 1 - F(-x)$, and if, for $a, b \in S$, $P_{ab} = F(w_a - w_b)$.

The definition of a linear model implies that Axiom 2.1 holds. Axiom 2.2 implies F must be continuous.

Brunk (1960) imterprets linear models as intrinsic worth models. He lets the w_a's denote the worths of the elements of S. Hence using these worths the elements of S can be ordered.

DEFINITION. F is the defining distribution of a linear model if $p_{ab} = F(w_a - w_b)$. The linear model is then said to be based on F. If F is absolutely continuous and f is the density of F then f is the defining density of the linear model, and the linear model is said to be based on the density f.

DEFINITION. $\mathcal{F}$ is the class of all cdf's F such that $F(x) = 1 - F(-x)$ and $F^{-1}(p)$ exists uniquely for $p \in (0, 1)$.

AXIOM 3.1. The set of probabilities $\{p_{ab}\}$ is generated by a linear model whose defining cdf is an element of $\mathcal{F}$.

When Axiom 3.1 holds then

$$\begin{aligned} p_{ac} &= F(w_a - w_c) \\ &= F((w_a - w_b) + (w_b - w_c)) \\ &= F\big(F^{-1}(p_{ab}) + F^{-1}(p_{bc})\big). \end{aligned}$$

Hence the composition rule is

$$G_F(p_1, p_2) = F\big(F^{-1}(p_1) + F^{-1}(p_2)\big).$$

DEFINITION. When Axiom 3.1 holds and the linear model is based on F, then $G_F(p_1, p_2)$ denotes the composition rule based on F.

The notation $G_F(p_1, p_2)$ will be taken to imply that Axiom 3.1 holds.

Axiom 3.1 is consistent with Axioms 2.1 and 2.2. Hence all composition rules generated by a linear model will satisfy Properties 2.1 through 2.4.

THEOREM 3.1. *Given a paired comparison model for which Axioms* 2.1 *and* 2.2 *hold, let G be the composition rule that this model generates. Then G satisfies Axioms* 2.3 *and* 2.4 *if and only if there exists a* cdf *F that satisfies Axiom* 3.1 *and defines a composition rule G_F that is identical to G.*

PROOF. First assume Axiom 3.1 holds. Then

$$G_F(p_1, p_2) = G_F(p_2, p_1)$$

and Axiom 2.3 holds.

Now also assume $p_1, p_2, p_3 \in (0, 1)$ and $G_F(p_1, p_2) \in (0, 1)$. Then

$$p_1 > p_3 \Leftrightarrow F^{-1}(p_1) > F^{-1}(p_3) \qquad \text{since} \quad F \in \mathcal{F};$$

$$\Leftrightarrow F^{-1}(p_1) + F^{-1}(p_2) > F^{-1}(p_3) + F^{-1}(p_2) \qquad \text{since} \quad |F^{-1}(p_2)| < \infty;$$

$$\Leftrightarrow F\big(F^{-1}(p_1) + F^{-1}(p_2)\big) > F\big(F^{-1}(p_3) + F^{-1}(p_2)\big) \qquad \text{since} \quad G_F(p_1, p_2) \in (0, 1).$$

Therefore the second part of Axiom 2.4 holds. The proof that the first part also holds is similar except that the possibility that $G_F(p_1, p_2) = 0$ or 1 rules out the strict inequality and the reverse implication.

Now assume Axioms 2.3 and 2.4 hold. Then Properties 2.5 through 2.8 can be applied.

Following Debreu (1958) it can be shown that there exists a continuous strictly increasing transformation T of $[\frac{1}{2}, 1)$ into the reals such that $T(\frac{1}{2}) = 0$ and the isoprobability lines transform into parallel straight lines. The existence of such a transformation T is a particular case of a problem of plane topology. The problem can be roughly described as follows: given three families of curves in a plane, when does there exist a topological transformation carrying them into three families of parallel straight lines? In Figure 3.1 the three families are the isoprobability lines, the vertical axis and the horizontal axis. After the transformation they are the lines with slope of minus one, the vertical axis and the horizontal axis.

Since T is continuous and strictly increasing on $[\frac{1}{2}, 1)$, then T^{-1} exists and is also continuous and strictly increasing on $[0, T(1)]$ where $T(1) = \lim_{p \to 1} T(p)$. Define F by

$$\begin{aligned} F(x) &= T^{-1}(x) \qquad \text{if} \quad x \in [0, T(1)) \\ &= 1 \qquad \text{if} \quad x \geqslant T(1) \\ &= 1 - F(-x) \qquad \text{if} \quad x < 0. \end{aligned}$$

Then F is a cdf and is an element of $\mathcal{F}$ (i.e., a linear model can be based on F.) It remains to verify that the composition rule generated by F is the same as the original one. Let $p_1 \in (\frac{1}{2}, 1)$ and $p_2 \in (\frac{1}{2}, 1)$. Assume $G(p_1, p_2) = p_3 \in (\frac{1}{2}, 1)$. Then $G(p_1, p_2) = G(p_3, \frac{1}{2})$ implies $T(p_1) + T(p_2) = T(p_3) + T(.5)$, since T maps the isoprobability line I_{p_3} into a straight line with slope of minus one. Now $T(\frac{1}{2}) = 0$.

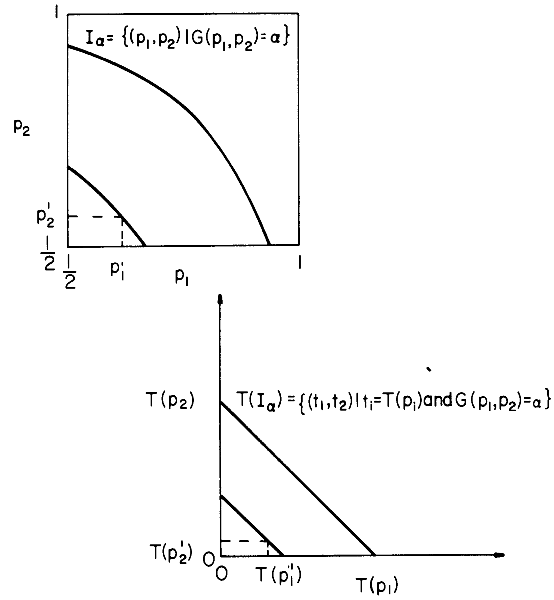


FIG. 3.1. Effect of T on the isoprobability lines.

Therefore $F(F^{-1}(p_1) + F^{-1}(p_2)) = F(T(p_3)) = F(F^{-1}(p_3)) = p_3$. The case for $p_3 = 1$ is handled similarly.

It follows by Property 2.8 that since G_F and G are identical on $(\frac{1}{2}, 1) \times (\frac{1}{2}, 1)$ then they are identical on $(0, 1) \times (0, 1)$. □

The following property was motivated by Noether's 1960 paper in which he gave a procedure for estimating the worths of a finite set of elements $T \subset S$ using the results of a set of paired comparisons involving the elements of T and assuming a given linear model. Using these estimated worths, p_{ab} can be estimated for all $a, b \in T$. He noted that if the defining cdf $F(x)$ is replaced by $F(cx)$ for some $c > 0$ then the estimate of w_a is changed to $(1/c)w_a$ but the resulting estimate of p_{ab} does not change.

PROPERTY 3.1. *The composition rules based on F_1 and F_2 are identical if and only if $\exists c > 0$ such that $F_1(x) \equiv F_2(cx)$ for all $x \in (-\infty, \infty)$*

PROOF. Assume $F_1(x) = F_2(cx)$. Then clearly

$$G_{F_1}(p_1, p_2) = G_{F_2}(p_1, p_2) \quad \text{for all} \quad (p_1, p_2) \in (0, 1) \times (0, 1).$$

Now assume $G_{F_1}(p_1, p_2) \equiv G_{F_2}(p_1, p_2)$. By Property 2.5 F_1 and F_2 are continuous. Hence for all x where $F_1(x) \in (0, 1)$, $\exists c_x > 0$ such that $F_1(x) = F_2(c_x x)$. It is sufficient to show that c_x is a constant independent of x. Note $F_1(x) = 1 - F_1(-x)$ and $F_2(x) = 1 - F_2(-x)$, hence $c_x = c_{-x}$. Therefore it is sufficient to show that $c_x = c$ for all $x > 0$.

Let t be any positive real such that $F_1(t) \in (\frac{1}{2}, 1)$. Let $p_t = F_1(t) = F_2(c_t t)$. $G_{F_1}(p_t, p_t) = G_{F_2}(p_t, p_t)$ by assumption. Therefore

$$F_1(2F_1^{-1}(p_t)) = F_2(2F_2^{-1}(p_t))$$
$$F_1(2t) = F_2(2c_t t).$$

Hence $c_{2t} = c_t$.

Similarly, by an induction argument, it follows that $c_{lt} = c_t$ for all positive integers l.

Let $H_1(p) = G_{F_1}(p, p)$. Then $H_1(p)$ is a continuous function mapping (0, 1) onto (0, 1). Hence $\exists p \in (\frac{1}{2}, p_t)$ such that $H_1(p) = p_t$. Therefore

$$G_{F_1}(p, p) = p_t = G_{F_2}(p, p)$$
$$F_1(2F_1^{-1}(p)) = p_t = F_2(2F_2^{-1}(p)).$$

Hence

$$F_1^{-1}(p) = \tfrac{1}{2}F_1^{-1}(p_t) = \tfrac{1}{2}t$$
$$F_2^{-1}(p) = \tfrac{1}{2}F_2^{-1}(p_t) = \tfrac{1}{2}c_t t.$$

Therefore

$$F_1\left(\tfrac{1}{2}t\right) = F_2\left(\tfrac{1}{2}c_t t\right)$$

and

$$\frac{c_1}{2}t = c_t.$$

It follows by an induction argument that for every positive integer k, $c_{v(t,k)} = c_t$ where $v(t, k) = t2^{-k}$.

Combining these two results, it follows that $c_{v(t,k,l)} = c_t$ where $v(t, k, l) = tl2^{-k}$ for all positive integers k, l.

The set $\{tl2^{-k}\}k = 1, 2, \cdots, l = 1, 2, \cdots,$ is dense in the set of all positive reals. Hence by continuity $F_1(x) = F_2(c_t x)$ for all $x > 0$. □

The following property concerns the rate of change of $G_F(p_1, p_2)$. In particular if $\frac{1}{2} \leqslant p_2 \leqslant p_1 < 1$, then under certain conditions increasing p_1 will increase G_F as much or more than increasing p_2.

PROPERTY 3.2. *If F is absolutely continuous and has a unimodal density f, then $\frac{1}{2} > |\frac{1}{2} - p_1'| \geqslant |\frac{1}{2} - p_2'|$ and $G_F(p_1', p_2') \in (0, 1)$ imply*

$$\frac{\partial}{\partial p_1} G_F(p_1, p_2)|_{(p_1, p_2) = (p_1', p_2')} \geqslant \frac{\partial}{\partial p_2} G_F(p_1, p_2)|_{(p_1, p_2) = (p_1', p_2')}.$$

PROOF.

$$G_F(p_1, p_2) = F(F^{-1}(p_1) + F^{-1}(p_2))$$
$$\frac{\partial G_F(p_1, p_2)}{\partial p_i} = \frac{\partial F^{-1}(p_i)}{\partial p_i} f(F^{-1}(p_1) + F^{-1}(p_2)).$$

Let

$$F^{-1}(p_i') = x_i$$

then

$$\left.\frac{\partial F^{-1}(p_i)}{\partial p_i}\right|_{p_i = p_i'} = \frac{1}{f(x_i)}.$$

$F(x) = 1 - F(-x)$ for all $x \in (-\infty, \infty)$ implies $f(x) = f(-x)$. Therefore zero is a mode of f and

$$|x'| > |x''| \Rightarrow f(x') \leqslant f(x'').$$

Hence

$$|\tfrac{1}{2} - p_1'| \geqslant |\tfrac{1}{2} - p_2'| \Rightarrow |x_1| \geqslant |x_2|$$

and

$$f(x_1) \leqslant f(x_2).$$

Therefore

$$\left.\frac{\partial G_F(p_1, p_2)}{\partial p_1}\right|_{(p_1,p_2)=(p_1',p_2')} \geqslant \left.\frac{\partial G_F(p_1, p_2)}{\partial p_2}\right|_{(p_1,p_2)=(p_1',p_2')} \qquad \square$$

4. Convolution models. One of the most popular of the early paired comparison models is Thurstone's Case V (1927). In this model it is assumed that each element a involved in a paired comparison generates a sensation $X_a = Y_a + w_a$. The Y_a's are assumed to be independently distributed as $N(0, \sigma^2)$ random variables. The variance σ^2 is assumed to be the same for all elements of s. The element that produces the largest sensation is the one that is chosen. Hence

$$P_{ab} = \Phi\left([w_a - w_b][\sigma 2^{\frac{1}{2}}]^{-1}\right).$$

Using a scale change this becomes

$$P_{ab} = \Phi(w_a - w_b).$$

Mosteller (1951) derived the same form for p_{ab} when the sensations are correlated with each other by a constant correlation coefficient. As a result of this and other contributions by Mosteller, the model $p_{ab} = \Phi(w_a - w_b)$ is now commonly known as the Thurstone-Mosteller model.

If Y_a and Y_b are independent and identically distributed according to some absolutely continuous cdf that is not necessarily a normal cdf, then the resulting special class of linear models can be considered as generalizations of Thurstone's case V. Here

$$\begin{aligned} p_{ab} &= \operatorname{Prob}(X_a > X_b) \\ &= F(w_a - w_b) \end{aligned}$$

Table 4.1

Name	Sensation distribution	Defining distribution	Composition rule
Thurstone-Mosteller	Normal	Normal	$P_{ac} = \Phi[\Phi^{-1}(P_{ab}) + \Phi^{-1}(P_{bc})]$
Bradley-Terry	Extreme Value	Logistic	$P_{ac} = \dfrac{P_{ab}P_{bc}}{P_{ab}P_{bc} + (1 - P_{ab})(1 - P_{bc})}$
Cauchy	Cauchy	Cauchy	$P_{ac} = \dfrac{1}{\pi}\{\text{arc tan}[\tan(\pi P_{ab} - \frac{\pi}{2}) + \tan(\pi P_{bc} - \frac{\pi}{2})] + \frac{\pi}{2}\}$
			$P_{ac} = 1 \quad \text{if} \quad (2(1 - P_{ab}))^{\frac{1}{2}} + (2(1 - P_{bc}))^{\frac{1}{2}} \leqslant 1;$
Triangle	Uniform	Triangle	$= P_{ab} + P_{bc} - \frac{3}{2} + (2(1 - P_{ab}))^{\frac{1}{2}} + (2(1 - P_{bc}))^{\frac{1}{2}} - 2((1 - P_{ab})(1 - P_{bc}))^{\frac{1}{2}}$ $\text{if} \quad P_{ab} \geqslant \frac{1}{2}, P_{bc} \geqslant \frac{1}{2} \quad \text{and} \quad (2(1 - P_{ab}))^{\frac{1}{2}} + (2(1 - P_{bc}))^{\frac{1}{2}} > 1;$ $= P_{ab} - P_{bc} - \frac{1}{2} + (2P_{bc})^{\frac{1}{2}} - (2(1 - P_{ab}))^{\frac{1}{2}} + 2(P_{bc}(1 - P_{ab}))^{\frac{1}{2}}$ $\text{if} \quad 1 > P_{ab} \geqslant 1 - P_{bc} \geqslant \frac{1}{2}.$
Double Exponential	Exponential	Double Exponential	$P_{ac} = 1 - 2(1 - P_{ab})(1 - P_{bc}) \quad \text{if} \quad P_{ab} \geqslant \frac{1}{2}, P_{bc} \geqslant \frac{1}{2};$ $= 1 - \frac{1}{2}(\frac{1 - P_{ab}}{P_{bc}}) \quad \text{if} \quad P_{ab} \geqslant 1 - P_{bc} \geqslant \frac{1}{2}.$

where F is the cdf of $Y_b - Y_a$. In particular if H is the cdf of Y_a and Y_b then F is the convolution of $H(x)$ and $1 - H(-x)$. For this reason these models are denoted as convolution models and H is denoted as the sensation distribution of the model.

The assumption that H is absolutely continuous implies that F is also absolutely continuous. Hence it can easily be shown that $F(x) = 1 - F(-x)$. The additional assumption that $F^{-1}(p)$ exists uniquely for $p \in (0, 1)$ insures that $F \in \mathcal{F}$, the class of cdf's that generate composition rules. All of the convolution models that will be considered in this paper will have defining distributions F that are elements of $\mathcal{F}$.

The Bradley-Terry model in which $p_{ab} = \lambda_a/(\lambda_a + \lambda_b)$ is now also in common use. It was first proposed by Zermelo in 1929 and was reintroduced and popularized by Bradley and Terry in 1952. Ford also independently proposed it in 1957. Bradley (1953) derived $p_{ab} = \lambda_a/(\lambda_a + \lambda_b) = e^{w_a}/(e^{w_a} + e^{w_b}) = 1/(1 + e^{-(w_a - w_b)})$ from a linear model based on the "squared hyperbolic secant" or logistic density. Davidson (1969) demonstrated that the Bradley-Terry model can be expressed as a convolution model using the extreme value distribution as the sensation distribution. The composition rule that the Bradley-Terry model generates is $p_{ac} = p_{ab}\,p_{bc}/(p_{ab}\,p_{bc} + (1 - p_{ab})(1 - p_{bc}))$ (Luce, 1959).

Let $V_{ab} = 2(p_{ab} - 1/2)$. In terms of the V's the Bradley-Terry composition rule is $V_{ac} = (V_{ab} + V_{bc})/(1 + V_{ab}V_{bc})$. In a personal communication T. Cover noted that this is the Lorentz transformation for velocities.

Table 4.1 displays examples of convolution models along with their sensation distributions and defining distributions. The composition rule for the triangle model and the double exponential model is only partially defined. The remaining parts of the composition rule can be constructed using Property 2.1. By Property 3.1 the listed defining distribution is unique up to a scale constant. The listed sensation distribution is not unique.

5. The uniform model. The uniform distribution yields an interesting linear model that is not a convolution model. The defining cdf is the $U(-1, +1)$ distribution, and the composition rule is as follows:

$$\begin{aligned} G_U(p_1, p_2) &= p_1 + p_2 - \tfrac{1}{2} \quad \text{if} \quad 0 < p_1 + p_2 - \tfrac{1}{2} < 1 \\ &= 0 \quad \text{if} \quad p_1 + p_2 - \tfrac{1}{2} \leqslant 0 \\ &= 1 \quad \text{if} \quad p_1 + p_2 - \tfrac{1}{2} \geqslant 1. \end{aligned}$$

A proof that two i.i.d. random variables cannot have a $U(-1, +1)$ distribution is contained in a paper by M. L. Puri and Sen (1968). It then follows that the uniform model cannot be a convolution model.

Make the transformation $Q_i = p_i - \frac{1}{2}$ and define $G_F^Q(Q_1, Q_2) = G_F(p_1, p_2) - \frac{1}{2}$ (i.e., if $Q_{ab} = p_{ab} - \frac{1}{2}$ and $Q_{bc} = p_{bc} - \frac{1}{2}$ then $Q_{ac} = p_{ac} - \frac{1}{2} = G_F(p_{ab}, p_{bc}) - \frac{1}{2} = G_F^Q(Q_{ab}, Q_{bc})$.) Using the Q_i values and G_F^Q composition rule rather than the p_i

values and the G_F composition rule, the uniform model has the following additive form for its composition rule.

$$\begin{aligned} G_U^Q(Q_1, Q_2) &= Q_1 + Q_2 \quad &\text{if} \quad -\tfrac{1}{2} \leqslant Q_1 + Q_2 \leqslant \tfrac{1}{2} \\ &= -\tfrac{1}{2} \quad &\text{if} \quad -\tfrac{1}{2} \geqslant Q_1 + Q_2 \\ &= +\tfrac{1}{2} \quad &\text{if} \quad \tfrac{1}{2} \leqslant Q_1 + Q_2. \end{aligned}$$

This simple additive form for $G_F^Q(Q_1, Q_2)$ holds approximately for many other linear models when Q_1 and Q_2 are near zero. Table 5.1 illustrates this numerically. A more precise statement of the above fact is the following proposition.

PROPOSITION 5.1. *If a linear model has a defining* cdf $F(x)$ *whose density* $f(x)$ *is continuous, finite and nonzero at* $x = 0$, *then for all* $\alpha > 0$,

$$\frac{G_F^Q(Q_1, \alpha Q_1)}{Q_1 + \alpha Q_1} \to 1 \quad \text{as} \quad Q_1 \to 0$$

(i.e., *for* p_1, p_2 *near* $\frac{1}{2}$ *the composition rule behaves like that of the uniform model or like any other composition rule whose* cdf *satisfies the above conditions.*)

PROOF.

$$\begin{aligned} \frac{d}{dQ_1} G_F^Q(Q_1, \alpha Q_1) &= \frac{d}{dQ_1} F\left(F^{-1}\left(Q_1 + \tfrac{1}{2}\right) + F^{-1}\left(\alpha Q_1 + \tfrac{1}{2}\right)\right) \\ &= f\left(F^{-1}\left(Q_1 + \tfrac{1}{2}\right) + F^{-1}\left(\alpha Q_1 + \tfrac{1}{2}\right)\right) \\ &\quad \times \left\{ \frac{1}{f\left(F^{-1}\left(Q_1 + \tfrac{1}{2}\right)\right)} + \frac{\alpha}{f\left(F^{-1}\left(\alpha Q_1 + \tfrac{1}{2}\right)\right)} \right\} \end{aligned}$$

$$\frac{d}{dQ_1}(Q_1 + \alpha Q_1) = 1 + \alpha.$$

Now as $Q_1 \to 0$

$$\frac{\dfrac{d}{dQ_1} G_F^Q(Q_1, \alpha Q_1)}{\dfrac{d}{dQ_1}(Q_1 + \alpha Q_1)} \to \frac{f(0)\left[\dfrac{1}{f(0)} + \dfrac{\alpha}{f(0)}\right]}{1 + \alpha} = 1.$$

Therefore by L'Hospital's rule

$$\frac{G_F^Q(Q_1, \alpha Q_1)}{Q_1 + \alpha Q_1} \to 1 \quad \text{as} \quad Q_1 \to 0. \qquad \square$$

COROLLARY 5.1. *If a convolution model has a sensation distribution whose density is square summable then* $G_F^Q(Q_1, \alpha Q_1)/(Q_1 + \alpha Q_1) \to 1$ *as* $Q_1 \to 0$.

PROOF. Let $h(x)$ be the density of the sensation distribution. Then $f(x) = \int_{-\infty}^{\infty} h(t)h(x + t)\, dt$ is the density of the defining distribution. It follows that $f(x)$ is continuous, finite and nonzero at $x = 0$ since h is square summable. $\square$

NOTE. All of the examples in Section 4 have sensation distributions whose densities are square summable. Hence Corollary 5.1 applies to all of them. An example of a linear model where Proposition 5.1 does not hold follows.

EXAMPLE 5.1. Let

$$f(x) = \frac{1}{4(|x|)^{\frac{1}{2}}} \qquad -1 \leqslant x \leqslant 1 \quad x \neq 0$$
$$= 0 \qquad \text{otherwise,}$$
$$F(x) = \tfrac{1}{2} + \tfrac{1}{2}x^{\frac{1}{2}} \qquad 0 \leqslant x \leqslant 1$$
$$= \tfrac{1}{2} - \tfrac{1}{2}x^{\frac{1}{2}} \qquad -1 \leqslant x \leqslant 0;$$
$$F^{-1}(p) = 4\left(p - \tfrac{1}{2}\right)^2 \qquad p \geqslant \tfrac{1}{2}$$
$$= -4\left(p - \tfrac{1}{2}\right)^2 \qquad p < \tfrac{1}{2}.$$

Therefore if $Q_1 \geqslant 0$ and $Q_2 \geqslant 0$, then

$$G_F^Q(Q_1, Q_2) = \tfrac{1}{2}(4Q_1^2 + 4Q_2^2)^{\frac{1}{2}} = (Q_1^2 + Q_2^2)^{\frac{1}{2}}.$$

Let $\alpha = 1$; then $G_F^Q(Q_1, Q_1) = Q_1 2^{\frac{1}{2}}$. Therefore $G_F^Q(Q_l, Q_l)/(Q_l + Q_l) \to 1/2^{-\frac{1}{2}} \neq 1$.

Table 5.1 gives some idea of how quickly different linear models approach the uniform model as p_1 and p_2 approach $\frac{1}{2}$.

Table 5.1.

Model	G_F (.51, .51)	G_F (.55, .55)	G_F (.6, .6)
Uniform	.52	.6	.7
Thurstone-Mosteller	.519994	.59922	.69382
Bradley-Terry	.519992	.59901	.69231
Cauchy	.519980	.59765	.68343
Triangle	.519899	.59737	.68885
Double-Exponential	.5198	.595	.68
Example 5.1	.514142	.57071	.64142

6. Table of values for examples of linear models. This section consists mainly of a brief table giving the values of composition rules for various (p_1, p_2) values. The table lists values for the five composition rules listed as examples in Section 4 and the one generated by the uniform model. The values of (p_1, p_2) used are $p_i =$ .1, .2, .3, .4, .6, .7, .8 and 9. for $i = 1, 2$. Half of the entries in the tables appear to be missing, but they can be computed using Axiom 2.2.: $(G(p_1, p_2) = G(p_2, p_1))$.

7. A partial ordering on the linear models. The values of $G_F(p_1, p_2)$ induce a partial ordering on the linear models for paired comparisons. For example if

Table 6.1.
$G_F\ (p_1, p_2)$

Model	$p_1 \backslash p_2$	.6	.7	.8	.9
A: Uniform		.2	.3	.4	.5
B: Thurstone-Mosteller		.15192	.22448	.32999	.5
C: Bradley-Terry	.1	.14286	.20588	.30769	.5
D: Cauchy		.11091	.12801	.16915	.5
E: Triangle		.15279	.22621	.33192	.5
F: Double Exponential		.125	.16667	.25	.5
A		.3	.4	.5	.6
B		.27817	.37554	.5	.67001
C		.27273	.36842	.5	.69231
D	.2	.24202	.31657	.5	.83085
E		.27234	.38796	.5	.66808
F		.25	.33333	.5	.75
A		.4	.5	.6	.7
B		.39317	.5	.62446	.77552
C	.3	.39130	.5	.63158	.79412
D		.37844	.5	.68343	.87199
E		.38735	.5	.63204	.77379
F		.375	.5	.66667	.83333
A		.5	.6	.7	.8
B		.5	.60683	.72183	.84807
C	.4	.5	.60870	.72727	.85714
D		.5	.62156	.75798	.88909
E		.5	.61265	.72766	.84721
F		.5	.625	.75	.875
A		.7	.8	.9	1.0
B		.69382	.78164	.86323	.93760
C	.6	.69231	.77778	.85714	.93103
D		.68343	.75798	.83085	.90901
E		.68885	.77620	.86120	.94164
F		.68	.76	.84	.92
A		.8	.9	1.0	1.0
B		.78164	.85287	.91403	.96454
C		.77778	.84483	.90323	.95455
D	.7	.75798	.80814	.85871	.91818
E		.77620	.84919	.91715	.97540
F		.76	.82	.88	.94

A		.9	1.0	1.0	1.0
B		.86323	.91403	.95384	.98313
C		.85714	.90323	.94118	.97297
D	.8	.83085	.85871	.88909	.92970
E		.86120	.91715	.96491	.99683
F		.84	.88	.92	.96
A		1.0	1.0	1.0	1.0
B		.93760	.96454	.98313	.99481
C	.9	.93103	.95455	.97297	.98780
D		.90901	.91818	.92970	.94874
E		.94164	.97540	.99683	1.0
F		.92	.94	.96	.98

$G_{F_1}(.7, .8) = .95$ and $G_{F_2}(.7, .8) = .9$, then the linear model based on F_1 can be said to be more "extreme" at $(p_1, p_2) = (.7, .8)$ than the linear model based on F_2. The following definition extends this idea.

DEFINITION. A linear model based on F_1 is *more extreme than* a linear model based on F_2 if $G_{F_1}(p_1, p_2) \geqslant G_{F_2}(p_1, p_2)$ for all $(p_1, p_2) \in [(\frac{1}{2}, 1) \times (\frac{1}{2}, 1)]$ and is *strictly more extreme than* a linear model based on F_2 if in addition $G_{F_1}(p_1', p_2') > G_{F_2}(p_1', p_2')$ for some $(p_1', p_2') \in [(\frac{1}{2}, 1) \times (\frac{1}{2}, 1)]$.

The above definition can be applied to any composition rule, but in light of Theorem 3.1, the rest of the paper will treat only linear models.

DEFINITION. Let $G_{F_1} \gg G_{F_2}$, $G_{F_1} \underline{\gg} G_{F_2}$, $G_{F_1} = G_{F_2}$, $G_{F_1} \underline{\ll} G_{F_2}$, $G_{F_1} \ll G_{F_2}$ denote G_{F_1} is strictly more extreme than, more extreme than, identical to, less extreme than, strictly less extreme than G_{F_2} respectively.

PROPOSITION 7.1. *If* $G_{F_a} \gg G_{F_b}$ *then*

(a) $(p_1, p_2) \in [(0, \frac{1}{2}) \times (0, \frac{1}{2})] \Rightarrow G_{F_a}(p_1, p_2) \leqslant G_{F_b}(p_1, p_2)$
(b) $1 > p_1 \geqslant 1 - p_2 \geqslant \frac{1}{2} \Rightarrow G_{F_a}(p_1, p_2) \leqslant G_{F_b}(p_1, p_2)$
(c) $1 > p_2 \geqslant 1 - p_1 \geqslant \frac{1}{2} \Rightarrow G_{F_a}(p_1, p_2) \leqslant G_{F_b}(p_1, p_2)$
(d) $0 < p_2 \leqslant 1 - p_1 \leqslant \frac{1}{2} \Rightarrow G_{F_a}(p_1, p_2) \geqslant G_{F_b}(p_1, p_2)$
(e) $0 < p_1 \leqslant 1 - p_2 \leqslant \frac{1}{2} \Rightarrow G_{F_a}(p_1, p_2) \geqslant G_{F_b}(p_1, p_2)$.

PROOF.

(a) $G_{F_a}(p_1, p_2) = 1 - G_{F_a}(1 - p_2, 1 - p_1)$ by Property 2.1

$$\leqslant 1 - G_{F_b}(1 - p_2, 1 - p_1)$$
$$= G_{F_b}(p_1, p_2);$$

(b) $G_{F_a}(p_1, p_2) = q_a \geqslant \frac{1}{2}$ and $G_{F_b}(p_1, p_2) = q_b \geqslant \frac{1}{2}$ by Property 2.7. $G_{F_a}(q_a, 1 - p_2) = p_1 = G_{F_b}(q_b, 1 - p_2)$ by Property 2.3. Note that $G_{F_a}(p, 1 - p_2) \geqslant G_{F_b}(p, 1 - p_2)$ for all $p \geqslant \frac{1}{2}$. Hence Axiom 2.4 implies that $q_a \leqslant q_b$ or $G_{F_a}(p_1, p_2) \leqslant G_{F_b}(p_1, p_2)$.

(c) Follows from (b) by Axiom 2.3.
(d) Follows from (c) by Property 2.1.
(e) Follows from (d) by Axiom 2.3. □

Assume $G_{F_a} \gg G_{F_b}$; then Figure 7.1 illustrates the typical relation between $G_{F_a}(p_1, p_2)$ and $G_{F_b}(p_1, p_2)$. The relationships in the figure follow from Property 2.6, Property 2.7 and Proposition 7.1.

Examination of Proposition 7.1 and Figure 7.1 reveals that if p_1 and p_2 are both elements of $(0, \frac{1}{2})$ or both elements of $(\frac{1}{2}, 1)$, then the more extreme linear model results in a composition probability that is closer to the extreme value of 0 or 1. The relation reverses itself when $p_1 \in (0, \frac{1}{2})$ and $p_2 \in (\frac{1}{2}, 1)$. Hence the extreme ordering lives up to its name only when (p_1, p_2) is in either the upper right or lower left quadrant.

The "extreme" ordering is only a partial ordering as can be seen as follows:
Let F_a be the double exponential distribution.
Let F_b be the Cauchy distribution.
Then

$$G_{F_a}(.6, .6) = .68 \quad \text{and} \quad G_{F_a}(.8, .8) = .92$$
$$G_{F_b}(.6, .6) = .68343 \quad \text{and} \quad G_{F_b}(.8, .8) = .88909.$$

Hence neither is more extreme than the other.

An example of the extreme ordering is contained in the following proposition.

PROPOSITION 7.2. *The Bradley-Terry model is more extreme than the double exponential model.*

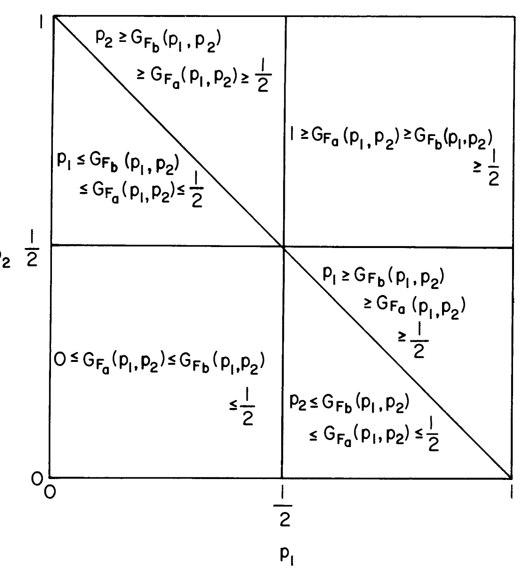


FIG. 7.1. Relation between $G_{F_a}(p_1, p_2)$ and $G_{F_b}(p_1, p_2)$ assuming $G_{F_a} \gg G_{F_b}$.

PROOF. Let

$$g(p_1, p_2) = G_{B-T}(p_1, p_2) - G_{\text{D.E.}}(p_1, p_2)$$
$$= \frac{p_1 p_2}{p_1 p_2 + (1-p_1)(1-p_2)} - [1 - 2(1-p_1)(1-p_2)]$$
$$\text{for } (p_1, p_2) \in \left[\left(\tfrac{1}{2}, 1\right) \times \left(\tfrac{1}{2}, 1\right)\right]$$

$$g(p_1, p_2)[p_1 p_2 + (1-p_1)(1-p_2)] = p_1 p_2 - [p_1 p_2 - 2p_1 p_2(1-p_1)(1-p_2) + (1-p_1)(1-p_2) - 2(1-p_1)^2(1-p_2)^2]$$
$$= 2p_1 p_2(1-p_1)(1-p_2) - (1-p_1)(1-p_2) + 2(1-p_1)^2(1-p_2)^2$$
$$= (1-p_1)(1-p_2)[2p_1 p_2 - 1 + 2(1-p_1)(1-p_2)]$$
$$= (1-p_1)(1-p_2)(2p_1 - 1)(2p_2 - 1)$$
$$> 0 \text{ for } (p_1 p_2) \in \left[\left(\tfrac{1}{2}, 1\right) \times \left(\tfrac{1}{2}, 1\right)\right]. \quad \square$$

8. The uniform model and the extreme partial ordering. THEOREM 8.1. *The uniform model is more extreme than any linear model that is based on a* cdf *F that has a unimodal density.*

PROOF. Let $p_1 \geqslant \frac{1}{2}$ and $p_2 \geqslant \frac{1}{2}$. By definition of a linear model, $F(x) = 1 - F(-x)$. Hence $f(x) = f(-x)$. Therefore f unimodal implies zero is a mode of f. Hence $f(t) \geqslant f(t+s)$ for $t \geqslant 0$, $s \geqslant 0$. Let $p_i = F(\Delta_i)$. Then $p_i \geqslant \frac{1}{2} \Rightarrow \Delta_i \geqslant 0$, $i = 1, 2$. Also

$$p_i = \tfrac{1}{2} + \textstyle\int_0^{\Delta_i} f(t)\, dt$$
$$G_F(p_1, p_2) = \tfrac{1}{2} + \textstyle\int_0^{\Delta_1 + \Delta_2} f(t)\, dt$$
$$G_F(p_1, p_2) - p_1 = \textstyle\int_{\Delta_1}^{\Delta_1 + \Delta_2} f(t)\, dt = \int_0^{\Delta_2} f(t + \Delta_1)\, dt$$
$$\leqslant \textstyle\int_0^{\Delta_2} f(t)\, dt$$
$$= p_2 - \tfrac{1}{2}.$$

Therefore $G_F(p_1, p_2) \leqslant p_1 + p_2 - \frac{1}{2}$. The result follows by comparing the above inequality with the uniform model composition rule. $\square$

COROLLARY 8.1. *The uniform model is more extreme than any convolution model which has a sensation distribution H which is absolutely continuous and unimodal.*

PROOF. Let F be the convolution of $H(x)$ and $1 - H(-x)$, then F will be unimodal [Hodges and Lehmann (1954)]. $\square$

In order for the uniform model to be more extreme than a linear model based on a cdf F it is not necessary that the density of F be unimodal, nor is it necessary that the sensation distribution of a convolution model have a unimodal density. The following example illustrates this.

EXAMPLE 8.1. Let the sensation distribution have the density

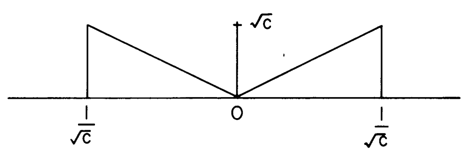
$$h(t) = \tfrac{1}{2} \qquad 0 < t < 1 \quad \text{or} \quad 2 < t < 3$$
$$= 0 \qquad \text{otherwise.}$$

It can easily be shown that the uniform model is more extreme than the linear model with the above sensation density.

An example of a linear model that is strictly more extreme than the uniform model will be necessarily based on a nonunimodal density. Such an example follows:

EXAMPLE 8.2. Let $f(x) = c|x|$ for $-c^{-\frac{1}{2}} \leqslant x \leqslant c^{-\frac{1}{2}}$. $c > 0$ is a scale factor. Then

$$F_{8.2}(x) = \tfrac{1}{2} + \tfrac{1}{2}cx^2 \qquad 0 \leqslant x \leqslant C^{-\frac{1}{2}}$$
$$= \tfrac{1}{2} - \tfrac{1}{2}cx^2 \qquad C^{-\frac{1}{2}} \leqslant x \leqslant 0.$$



COUNTEREXAMPLE. The composition rule $G_{F_{8.2}}$ based on the above distribution is strictly more extreme than the composition rule based on the uniform model.

9. The extreme partial ordering and the length of the tail of the defining distribution. In Theorem 8.1 and Proposition 7.2, the distribution with the "smaller" tail generates the more extreme linear model. This observation motivates the following definition for a partial ordering on the distributions in $\mathcal{F}$ according to the lengths of their tails.

DEFINITION. If $F_a, F_b \in \mathcal{F}$, then $F_a <_q F_b$ (F_a is q-ordered with respect to F_b) if and only if $G_{F_a} \gg G_{F_b}$ (the linear model generated by F_a is more extreme than the linear model generated by F_b).

Van Zwet (1964) also defined a weak ordering on the elements in $\mathcal{F}$ according to the lengths of their tails. He defined $F_1 <_s F_2$ if and only if $F_1(0) = F_2(0) = \frac{1}{2}$ and $F_2^{-1}(F_1(x))$ is concave-convex about the origin on the support of F_1, i.e., $\{x|0 < F_1(x) < 1\}$. Lawrence (1975) extended Van Zwet's partial ordering by defining $F_1 <_r F_2$ if and only if $F_1(0) = F_2(0) = \frac{1}{2}$ and $F_2^{-1}(F_1(x))/x$ is increasing (decreasing) for x positive (negative) on the support of F_1. Clearly $F_1 <_s F_2$ implies $F_1 <_r F_2$.

THEOREM 9.1. *Suppose $F_a, F_b \in \mathcal{F}$ and $F_a <_r F_b$, then $G_{F_a} \gg G_{F_b}$ and hence $F_a <_q F_b$.*

PROOF. Lawrence's definition implies that for any $c > 0$, there exists constants $M_1(c)$, $M_2(c)$ and $M_3(c)$ such that $0 \leqslant M_1(c) \leqslant M_2(c) \leqslant M_3(c) \leqslant \infty$ and

(a) $0 < x < M_1(c)$ implies $F_a(cx) > F_b(x)$;
(b) $M_1(c) \leqslant x \leqslant M_2(c)$ implies $F_a(cx) = F_b(x)$;
(c) $M_2(c) < x < M_3(c)$ implies $F_a(cx) < F_b(x)$;
(d) $x \geqslant M_3(c)$ implies $F_a(x) = F_b(x) = 1$.

Assume $1 > p_1 \geqslant p_2 > \frac{1}{2}$, let $F_b^{-1}(p_1) = \Delta_1$ and $F_b^{-1}(p_2) = \Delta_2$. By continuity there exists $d > 0$ such that $F_a(d\Delta_1) = p_1$. By (b) above $M_1(d) \leqslant \Delta_1 \leqslant M_2(d)$. Define $F_d(x) = F_a(dx)$. Let $F_d^{-1}(p_2) = \Delta_3$. Then $p_1 \geqslant p_2$ implies $\Delta_1 \geqslant \Delta_2$. By (a) above $F_d(x) \geqslant F_b(x)$ for $0 < x \leqslant \Delta_1$. Hence $F_d(\Delta_2) \geqslant F_b(\Delta_2)$. By definition $F_d(\Delta_3) = p_2 = F_b(\Delta_2)$. Therefore since F_d is strictly increasing on $\{x : F(x) \in (0, 1)\}$ it follows that $\Delta_2 \geqslant \Delta_3$. See Figure 9.1 for a graphical illustration. Hence

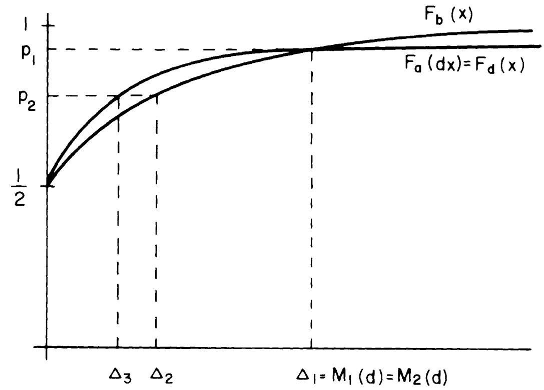


FIG. 9.1.

$$\begin{aligned} G_{F_d}(p_1, p_2) &= F_d(\Delta_1 + \Delta_3) \leqslant F_d(\Delta_1 + \Delta_2) \\ &\leqslant F_b(\Delta_1 + \Delta_2) \text{ by (c) and (d) above} \\ &= G_{F_b}(p_1, p_2). \end{aligned}$$

It follows by Property 3.1 that $G_{F_a}(p_1, p_2) \leqslant G_{F_b}(p_1, p_2)$. ▯

PROPOSITION 9.1. *The Thurstone-Mosteller model is more extreme than the Bradley-Terry model.*

PROOF. Van Zwet (1964) showed that if

$$F_a(x) = \frac{1}{1 + e^{-x}} \qquad -\infty < x < \infty$$

$$F_b(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt \qquad -\infty < x < \infty$$

then $F_b <_s F_a$. His proof only involves taking repeated derivatives of $F_a^{-1}(F_b(x))$ and showing it is convex. F_a is the logistic cdf and the Bradley-Terry model is based on it. $F_b(x)$ is the standard normal cdf and the Thurstone-Mosteller model is based on it. Also $F_b <_s F_a \Rightarrow F_b <_r F_a$. Hence by Theorem 9.1, the Thurstone-Mosteller model is more extreme than the Bradley-Terry model. ▯

Lawrence (1975) and Doksum (1969) both state that double exponential $_s<$ Cauchy which would imply that the double exponential model is more extreme than the Cauchy model. It has already been noted that the double exponential model and the Cauchy model cannot be ordered by the extreme partial ordering. Both Lawrence and Doksum cite van Zwet (1964) for proof of their statement. There is no mention of this s-ordering in van Zwet's paper. A simple direct proof that the ordering does not hold follows.

COUNTEREXAMPLE. If $F_a(x) = 1/\pi[\arc \tan x + (\pi/2)]$

$$F_b(x) = \tfrac{1}{2}e^{x} \qquad x < 0$$
$$= 1 - \tfrac{1}{2}e^{-x} \qquad x \geqslant 0.$$

Then $F_b(x) \not<_r F_a(x)$, $F_a(x) \not<_r F_b(x)$, $F_b(x) \not<_s F_a(x)$, and $F_a(x) \not<_s F_b(x)$.

PROOF OF COUNTEREXAMPLE.

$$F_a^{-1}[F_b(x)] = \tan\left\{\pi\left[1 - \tfrac{1}{2}e^{-x}\right] - \frac{\pi}{2}\right\} \qquad \text{for} \quad x > 0.$$

Let $x_0 = .6$.

λ	$F_a^{-1}[F_b(\lambda x_0)]$	$\lambda F_a^{-1}[F_b(x_0)]$
0	0	0
$\frac{1}{3}$	.29269	.28577
$\frac{2}{3}$	.56972	.57154
1	.85731	.85731

Clearly $F_a^{-1}[F_b(x)]$ and $F_b^{-1}[F_a(x)]$ are neither starshaped nor convex on $x > 0$. □

THEOREM 9.2. *Suppose*

(A) f_a *and* f_b *are densities whose associated* cdf's, F_a *and* F_b, *are elements of* $\mathfrak{F}$.

(B) *For every* $c > 0$ *there exists* $N_1(c) > N_2(c) \geqslant 0$ *such that*

(i) $|t| < N_2(c) \Rightarrow f_b(t) < cf_a(ct)$

(ii) $N_2(c) < |t| < N_1(c) \Rightarrow f_b(t) > cf_a(ct)$

(iii) $|t| > N_1(c) \Rightarrow f_b(t) < cf_a(ct)$.

Then the linear model based on f_b *is more extreme than the linear model based on* f_a.

PROOF. It is sufficient to show that for all $c > 0$ there exists constants $M_1(c)$, $M_2(c)$, and $M_3(c)$ such that conditions (a), (b), (c) and (d) of Theorem 9.1 are satisfied. See Figure 9.2 for a graphical illustration.

Note that $cf_a(cx)$ is the density of $F_a(cx)$.

Let $x \geqslant N_1(c)$. Then by Assumption B (iii)

$$\int_x^\infty f_b(t)\,dt < \int_x^\infty cf_a(ct)\,dt.$$

Therefore $1 - F_b(x) < 1 - F_a(cx)$ or $F_b(x) > F_a(cx)$.

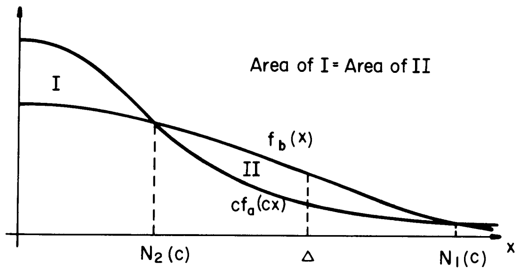


FIG. 9.2.

If $N_2(c) = 0$, then $0 \leqslant x < N_1(c)$ implies

$$\int_0^x f_b(t)\, dt > \int_0^x cf_a(ct)\, dt.$$

Therefore $F_b(x) - \frac{1}{2} > F_a(cx) - \frac{1}{2}$ or $F_b(x) > F_a(cx)$. Hence conditions (a), (b), (c) and (d) are satisfied with $M_1(c) = M_2(c) = 0$ and $M_3(c) = \infty$.

Now assume $N_2(c) > 0$ and let $0 < x < N_2(c)$. Then

$$\int_0^x f_b(t)\, dt < \int_0^x cf_a(ct)\, dt.$$

Therefore $F_b(x) < F_a(cx)$.

By continuity of F_a and F_b there exists a $\Delta > 0$ such that $F_a(c\Delta) = F_b(\Delta)$. By the above remarks $N_2(c) < \Delta < N_1(c)$.

Now let $N_2(c) < x < \Delta$. Then by Assumption B (ii)

$$\int_x^\Delta f_b(t)\, dt > \int_x^\Delta cf_a(ct)\, dt.$$

Therefore $F_b(\Delta) - F_b(x) > F_a(c\Delta) - F_a(cx)$ or $F_b(x) < F_a(cx)$.

Finally let $\Delta < x < N_1(c)$. Then by Assumption B (ii)

$$\int_\Delta^x f_b(t)\, dt > \int_\Delta^x cf_a(ct)\, dt.$$

Therefore $F_b(x) - F_b(\Delta) > F_a(cx) - F_a(c\Delta)$ or $F_b(x) > F_a(cx)$.

It follows that $0 < x < \Delta$ implies $F_a(cx) > F_b(x)$ and $x > \Delta$ implies $F_a(cx) < F_b(x)$. Hence conditions (a), (b), (c) and (d) of Theorem 9.1 are satisfied with $M_1(c) = M_2(c) = \Delta$ and $M_3(c) = \infty$. □

Theorem 9.2 can be made more general by use of nonstrict inequalities in the assumptions. The resulting proof will need more details. Using the nonstrict inequalities we can see that Theorem 9.2 can be thought of as a generalization of Theorem 8.1. In Theorem 8.1 f_b is the uniform density on $(-1, +1)$ and f_a is any unimodal density that is symmetric about zero. Then $cf_a(ct)$ and $f_b(t)$ will "intersect" ($cf_a(ct) \leqslant f_b(t)$ changes to $cf_a(ct) > f_b(t)$ or vice-versa) at most two times. Hence if nonstrict inequalities are used in Theorem 9.2 then this pair of densities (f_a and f_b) will satisfy the conditions.

The following tree summarizes some of the results concerning the extreme partial ordering

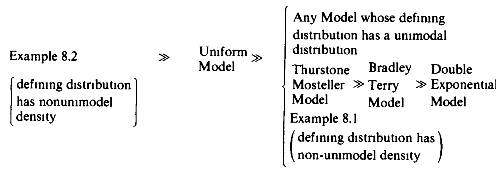


FIG. 9.3.

Acknowledgment. I am indebted to the referee for his constructive comments.

REFERENCES

[1] BRADLEY, R. A. (1953). Some statistical methods in taste testing and quality evaluation. *Biometrics* **9** 22–38.

[2] BRADLEY, RALPH A. (1976). Sciences, statistics, and paired comparison (with discussion). *Biometrics* **32** 213–239.

[3] BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs I, the method of paired comparisons. *Biometrika* **39** 324–345.

[4] BRUNK, H. D. (1960). Mathematical models for ranking from paired comparisons. *J. Amer. Statist. Assoc.* **55** 503–520.

[5] DAVID, H. A. (1963). *The Method of Paired Comparisons*. Griffin, London.

[6] DAVIDSON, R. R. (1969). On a relation between two representations of a model for paired comparisons. *Biometrics* **25** 597–599.

[7] DEBREU, G. (1958). Stochastic choice and cardinal utility. *Econometrica* **26** 440–444.

[8] DE CANI, J. S. (1969). Maximum likelihood paired comparison ranking by linear programming. *Biometrika* **56** 537–545.

[9] DE CANI, J. S. (1972). A branch and bound algorithm for maximum likelihood paired comparison ranking. *Biometrika* **59** 131–135.

[10] DOKSUM, K. (1969). Starshaped transformations and the power of rank tests. *Ann. Math. Statist.* **40** 1167–1176.

[11] FORD, L. R. (1957). Solution of a ranking problem from binary comparisons. *Amer. Math. Monthly* **64** 28–33.

[12] HODGES, J. L. and LEHMANN, E. L. (1954). Matching in-paired comparisons. *Ann. of Math. Statist.* **25** 777–791.

[13] LAWRENCE, M. J. (1975). Inequalities of s-ordered distributions. *Ann. Statist.* **3** 413–428.

[14] LUCE, R. D. (1959). *Individual Choice Behavior*. Wiley, New York.

[15] MOSTELLER, F. (1951). Remarks on the method of paired comparisons: I The least-square solution assuming equal standard deviations and equal correlations. II The effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. III A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika* **16** 3–9, 203–206, 207–218.

[16] NOETHER, G. F. (1960). Remarks about a paired comparison model. *Psychometrika* **25** 357–367.

[17] PURI, M. L. and SEN, P. K. (1968). On Chernoff-Savage tests for order alternatives in randomized blocks. *Ann. Math. Statist.* **39** 967–972.

[18] REMAGE, R. and THOMPSON, W. A. (1966). Maximum-likelihood paired comparison rankings. *Biometrika* **53** 143–149.

[19] THOMPSON, W. A. and REMAGE, R. (1964). Rankings from paired comparisons. *Ann. Math. Statist.* **35** 739–747.

[20] THURSTONE, L. L. (1927). A law of comparative judgement. *Psychological Review* **34** 273–286.

[21] VAN ZWET, W. R. (1964). *Convex Transformations of Random Variables*. Math. Centre, Amsterdam.

[22] ZERMELO, E. (1929). Die berechnung Turnier-Ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Math. Z.* **29** 436–460.

DEPARTMENT OF STATISTICS
UNIVERSITY OF KENTUCKY
LEXINGTON, KENTUCKY 40506