On how to decide which of two populations is best

Eindhoven University
of Technology

Department of Mathematics and Computing Sciences

Memorandum COSOR 98-10

On how to decide which of two populations is best

P. van der Laan
C. van Eeden

Eindhoven, May 1998
The Netherlands

ON HOW TO DECIDE WHICH OF TWO POPULATIONS IS BEST

Paul van der Laan
Eindhoven University of Technology
Eindhoven, The Netherlands

Constance van Eeden
University of British Columbia
Vancouver, Canada

Abstract

In this paper we consider the problem of deciding which of two populations π_1 and π_2 has the larger location parameter. We base this decision - which is a choice between "π_1", "π_2" and "π_1 or π_2" - on summary statistics X_1 and X_2, obtained from independent samples from the two populations. Our loss function contains a penalty for the absence of a "good" population as well as for the presence of a "bad" one among those chosen. We show that, for our class of decision rules (see (1.2)), the one that chooses the population with the largest observed value of X_i minimizes the expected loss. It also, obviously, minimizes the expected number of chosen populations. We give conditions under which the expected loss has a unique maximum and, for several examples where these conditions are satisfied, we also show that the expected loss is, for each (θ_1, θ_2), strictly decreasing in the (common) sample size n. For the case of normal populations Bechhofer (1954) proposed and studied this decision rule where he chose n to lower-bound the probability of a correct selection. Several new results on distributions having increasing failure rate, needed for our results, are of independent interest, as are new results on the peakedness of location estimators.

Keywords: *Decision theory; two-sample problem; selection; loss function; good populations; bad populations; location parameter; failure rate; peakedness*
1991 AMS Subject Classifications: *62F07; 62F11*
1998 STMA Subject Classifications: *05:130; 04:020*

1 Introduction

Consider two populations π_1 and π_2 and let X_1 and X_2 be independent summary statistics obtained from samples from π_1 and π_2 respectively, where X_i has distribution function $G(x-\theta_i), i = 1,2$, and G is known and continuous. The problem considered in this paper is one of deciding, on the basis of (X_1, X_2), which of the parameters θ_1 and θ_2 is the larger one, where we allow for the possibility of deciding that we do not know which one is larger. Our loss function contains a penalty for not including at least one "good" population among our candidates for the larger θ_i as well as a penalty for including a "bad" one. Here the population π_i is "good" (resp. "bad") when, for a given $\varepsilon \geq 0$, $\theta_{[2]} - \theta_i \leq \varepsilon$ (resp. $\theta_{[2]} - \theta_i > \varepsilon$) where $\theta_{[1]} < \theta_{[2]}$. In van der Laan and van Eeden (1998) a loss function is used where penalties are given only for losses due to the absence of good populations in the selected subset and not for losses due to the presence of bad ones. In the case where $\theta_{[2]} - \theta_{[1]} \leq \varepsilon$, both populations are good and we take the loss to be zero, no matter which decision is taken. In the case where $\theta_{[2]} - \theta_{[1]} > \varepsilon$, the penalty when choosing only θ_i with $\theta_i \neq \theta_{[2]}$ is $(\theta_{[2]} - \theta_i - \varepsilon)^p$, where p is a given positive constant. If the decision taken is that we do not know which is the larger one then the penalty for this case is $(\theta_{[2]} - \theta_{[1]} - \varepsilon)^p$. More formally, our loss function $L(\theta, d)$ is defined by

$$L(\theta,d) = \begin{cases} 0 \quad \text{for all } d & \text{when } \theta_{[2]} - \theta_{[1]} \leq \varepsilon; \\ \sum_{i=1}^{2} \left(\theta_{[2]} - \theta_i - \varepsilon\right)^p I(d = d_i) I(\theta_{[2]} - \theta_i > \varepsilon) & \\ + \left(\theta_{[2]} - \theta_{[1]} - \varepsilon\right)^p I(d = d_{12}) & \text{when } \theta_{[2]} - \theta_{[1]} > \varepsilon, \end{cases} \tag{1.1}$$

where $\theta = (\theta_1, \theta_2)$, d is the decision taken, $d_i, i = 1,2$, is the decision that θ_i is the larger of θ_1 and θ_2 and d_{12} is the decision that we do not know which of θ_1 and θ_2 is the larger one.

The selection rule δ_c is given by

$$\begin{aligned} d = d_i \quad & \text{when } (X_i = X_{[2]}, X_j < X_{[2]} - c, j = 1,2, j \neq i), i = 1,2, \\ d = d_{12} \quad & \text{when } |X_1 - X_2| \leq c, \end{aligned} \tag{1.2}$$

where $X_{[1]} \leq X_{[2]}$ are the ordered X_i's and c is a non-negative constant.

In Section 2 it will be shown that the rule with $c = 0$ minimizes, uniformly in $\mu = |\theta_2 - \theta_1|$, the expected loss. It obviously also minimizes the number of chosen populations, because $c = 0$ means that we decide that the population whose summary statistic has the largest observed value is the one with the largest θ_i. We also show there that, when the distribution of $X_1 - \theta_1 - (X_2 - \theta_2)$ has nondecreasing failure rate (IFR), the expected loss has a unique maximum which is attained. In Section 3 we give several examples of populations and summary statistics for which the condition of IFR is satisfied. In these examples it is also shown that the expected loss is strictly decreasing in the (common) sample size. Section 4 contains some auxiliary (known, but not easy to find in the literature) results needed for some of our results.

2 Some properties of the risk function

The following Theorem 2.1 gives the risk funcion $R(\theta, \delta_c, \varepsilon) = \mathcal{E}_\theta L(\theta, \delta_c(X))$ of the decision rule δ_c.

Theorem 2.1 *The risk function of the rule δ_c is given by*

$$R(\theta, \delta_c, \varepsilon) = \mathcal{E}_\theta L(\theta, \delta_c(X)) = (\mu - \varepsilon)^p P_\theta(Z_2 - Z_1 \leq c - \mu) I(\mu > \varepsilon), \tag{2.1}$$

where $Z_1 = X_1 - \theta_1$ and $Z_2 = X_2 - \theta_2$ are independent and identically distributed random variables with distribution function G and $\mu = |\theta_1 - \theta_2|$.

Proof. Assume without loss of generality that $\theta_2 > \theta_1$. Then

$$\begin{aligned}
R(\theta, \delta_c, \varepsilon) &= (\theta_2 - \theta_1 - \varepsilon)^p I(\mu > \varepsilon) P_\theta(-c < X_1 - X_2 < c) \\
&+ (\theta_2 - \theta_1 - \varepsilon)^p I(\mu > \varepsilon) P_\theta(X_1 - X_2 > c) \\
&= (\mu - \varepsilon)^p P_\theta(X_1 - X_2 > -c) I(\mu > \varepsilon) \\
&= (\mu - \varepsilon)^p P_\theta(X_2 - X_1 < c) I(\mu > \varepsilon) \\
&= (\mu - \varepsilon)^p P_\theta(Z_2 - Z_1 < c - (\theta_2 - \theta_1)) I(\mu > \varepsilon),
\end{aligned}$$

from which the result follows immediately. □

From Theorem 2.1 it folows that $R(\theta, \delta_c, \varepsilon)$ is, for each $\mu > \varepsilon$, nondecreasing in c. So, the rule δ_c with $c = 0$ minimizes, uniformly in μ, the expected loss. This rule δ_0 is given by

$$i \in d \Longleftrightarrow X_i \geq X_{[k]}, \tag{2.2}$$

which is equivalent to the rule

$$\left.\begin{array}{l}\text{select, as the best population, the one} \\ \text{which gives the largest observed value of } X_i.\end{array}\right\} \tag{2.3}$$

This rule is Bechhofer's selection rule (see e.g. Bechhofer (1954)). He considers the case of k ($k \geq 2$) normal populations with a 0-1 loss function, where the loss is zero if and only if the selected population is the one with the largest θ_i. For samples of equal sizes n, he chooses n in such a way that, for $\theta_{[k]} - \theta_{[k-1]} \geq \delta^*$, $P_\theta(\text{correct selection}) \geq P^*$ for given $\delta^* > 0$ and $k^{-1} < P^* < 1$.

We use the loss function (1.1) and want to choose the (common) sample size n such that, for all (θ_1, θ_2), $\mathcal{E}_\theta L(\theta, \delta_0) \leq R_o$ for a given $R_o > 0$. Whether this is possible depends upon the shape of the risk function as a function of $\mu = \theta_{[2]} - \theta_{[1]}$ and n.

Some properties of the risk function as a function of μ are given in Theorem 2.2, where we assume that the following Conditions A(1) and A(2) are satisfied.

A(1) The distribution of $Z_2 - Z_1$ has IFR.

Note that, under Condition A(1), the support of $Z_2 - Z_1$ is an interval, $[-a, a]$ say for some $a > 0$.

A(2) The distribution function H of $Z_2 - Z_1$ has a derivative h which is continuous on $(-a, a)$.

From Theorem 2.1 it is seen that the risk function of δ_o is zero for all μ when $\varepsilon \geq a$. This can also be seen directly by noting that

$$\mu > \varepsilon \Longrightarrow \mu > a \quad \text{when } \varepsilon \geq a$$

and that $\mu > a$ implies that

$$P_\theta(X_2 - X_1 < 0) = 1 \quad \text{for all } \theta \text{ with } \theta_2 - \theta_1 < -a,$$

$$P_\theta(X_2 - X_1 > 0) = 1 \quad \text{for all } \theta \text{ with } \theta_2 - \theta_1 > a.$$

So, when $\varepsilon \geq a$, the rule δ_o always selects the population with the largest location parameter. Given that G, and thus H, is known, a is known, so one knows whether or not the chosen ε satisfies $\varepsilon \geq a$. In what follows we suppose that $\varepsilon < a$.

Theorem 2.2 *Under the conditions A(1) and A(2), the risk function of the decision rule δ_o is strictly unimodal in μ.*

Proof. First note that (see Theorem 2.1) the risk function of δ_o is zero for $\mu \leq \varepsilon$ as well as for $\mu \geq a$. For $\varepsilon < \mu < a$

$$\frac{d}{d\mu} \log R(\theta, \delta_0, \varepsilon) = \frac{p}{\mu - \varepsilon} - \frac{h(-\mu)}{H(-\mu)} = \frac{p}{\mu - \varepsilon} - \frac{h(\mu)}{1 - H(\mu)} \,. \tag{2.4}$$

Now note that $1 - H(\mu) > 0$ for all $\mu \in [0, a)$ and that, by Condition A(2), $h(\mu) < \infty$ for all $\mu \in [0, a)$. So, $(d/d\mu) \log R(\theta, \delta_0, \varepsilon)$ is strictly decreasing in μ for $\mu \in (\varepsilon, a)$ with

$$\begin{aligned} \lim_{\mu \to \varepsilon} \frac{d}{d\mu} \log R(\theta, \delta_0, \varepsilon) &= \infty, \\ \limsup_{\mu \to a} \frac{d}{d\mu} \log R(\theta, \delta_0, \varepsilon) &= \frac{p}{a - \varepsilon} - \liminf_{\mu \to a} \frac{h(\mu)}{1 - H(\mu)} < 0. \end{aligned} \tag{2.5}$$

The inequality in the second line of (2.5) follows, for $a = \infty$, from the fact that $h(\mu)/(1 - H(\mu)) \geq h(0)/(1 - H(0)) > 0$ for all $\mu \in (-\infty, \infty)$. To prove the inequality for $a < \infty$, first note that, by Condition A(2),

$$- \log(1 - H(\mu)) = - \log(1 - H(\mu_o)) + \int_{\mu_o}^{\mu} \frac{h(t)}{1 - H(t)} dt \quad \text{for } - a < \mu_o < \mu < a. \tag{2.6}$$

The left hand side of (2.6) converges to infinity as $\mu \to a$, so the second term in its righthand side converges to infinity as $\mu \to a$, which implies, by Condition A(1), that $h(\mu)/(1 - H(\mu)) \to \infty$ as $\mu \to a$. Thus, by the continuity of $h(\mu)/(1 - H(\mu))$ for $-a < \mu < a$ (see Condition A(2)),

$$(d/d\mu) R(\theta, \delta_0, \varepsilon) = 0$$

has exactly one solution in $\mu \in (\varepsilon, a)$, which (together with 2.5) proves the result. $\square$

The next section contains some examples where the conditions A(1) and A(2) are satisfied and the risk function is, for each $\mu \in (\varepsilon, a)$, strictly decreasing in n. By Theorem 2.2 the sample size can, for such examples, be chosen such that $R(\theta, \delta_o, \varepsilon) \leq R_o$ for a given $R_o > 0$.

3 Examples

For each of the examples below we have chosen the summary statistics and the density f such that the Z_i have a symmetric distribution and we will show that, for each example, $Z_2 - Z_1$ has a nondecreasing failure rate (IFR), i.e. we show that

$$FR_h(x) = \frac{h(x)}{1 - H(x)} \tag{3.1}$$

is nondecreasing on $\{x | H(x) < 1\}$. We also show that, in each case, the risk function is, for each μ, strictly decreasing in the common sample size n.

For these proofs we need the notions of, and the relationships between, IFR, logconcavity of a density, Pólya frequency functions, strong unimodality of a distribution function and peakedness of a random variable. We have assembled what we need about this, with references to the relevant literature, in Section 4.

In the proofs of the IFR of $Z_2 - Z_1$ we use, several times, the fact that, if W_1 and W_2 are independent and each have IFR, then $W_1 + W_2$ has IFR. This result can be found e.g. in Barlow, Marshall and Proschan ((1963), p. 380).
Because in each of our examples the Z_i have a distribution which is symmetric around zero, $Z_2 - Z_1$ and $Z_2 + Z_1$ have the same distribution. So, by this Barlow-Marshall-Proschan result, we have

Lemma 3.1 *If the Z_i's have symmetric distributions then*

- *a) $Z_2 - Z_1$ has IFR when each of Z_1 and Z_2 has IFR;*
- *b) When the Z_i's are sample means $\sum_{j=1}^{n} X_{i,j}/n$, $Z_2 - Z_1$ has IFR when the $X_{i,j}$ have IFR.*

In the case where the summary statistic is the median of a sample of an odd number of observations, the IFR of X_i is a special case of the following result.

Lemma 3.2 *The k^{th} order statistic $Y_{k:n}$ of a sample $Y_1, \ldots, Y_n$ from a distribution with a Lebesgue density has, when n is odd, IFR when the Y_i have IFR.*

Proof. The distribution function of $Y_{k:n}$ is given by

$$P(Y_{k:n} \leq y) = \sum_{i=k}^{n} \binom{n}{i} (F(y))^i (1 - F(y))^{n-i},$$

where f and F are, respectively, the density and distribution function of the Y_i.
So,

$$1 - P(Y_{k:n} \leq y) = \sum_{i=0}^{k-1} \binom{n}{i} (F(y))^i (1 - F(y))^{n-i}.$$

Further, the density of $Y_{k:n}$ is given by

$$\frac{n!}{(k-1)!(n-k)!} f(y)(F(y))^{k-1}(1 - F(y))^{n-k}.$$

So, the failure rate of $Y_{k:n}$ is given by

$$K_n \frac{f(y)}{1 - F(y)} \frac{1}{\sum_{i=0}^{k-1} \binom{n}{i} \left((F(y))^{-1}(1 - F(y))\right)^{k-i-1}},$$

where K_n is a positive constant. Further, $f(y)/(1-F(y))$ is nondecreasing in y because Y_1 has IFR and

$$\left(F(y)^{-1}(1-F(y))\right)^{k-i-1}$$

is nonincreasing in y. □

The result proved in Lemma 3.2 is stated, without proof, in Szekli's (1995) problem D, p. 28.

For the IFR of the midrange the following result holds for a sample from a uniform distribution.

Lemma 3.3 *For a sample $Y_1, \ldots, Y_n$ from a uniform distribution on the interval $[-1,1]$, the midrange*

$$T = \frac{1}{2}(\min_{1\leq i\leq n} Y_i + \max_{1\leq i\leq n} Y_i)$$

has IFR.

Proof. The joint density of $\min_{1\leq i\leq n} Y_i$ and $\max_{1\leq i\leq n} Y_i$ at (x,y) is, for $n \geq 2$, given by

$$\frac{n(n-1)}{2^n}(y-x)^{n-2} \qquad -1 \leq x < y \leq 1.$$

So,

$$P(\min_{1\leq i\leq n} Y_i + \max_{1\leq i\leq n} Y_i \leq 2t) =$$

$$\frac{n(n-1)}{2^n}\int_{-1}^{t} dx \int_{x}^{2t-x} (y-x)^{n-2} dy = \frac{(1+t)^n}{2} \qquad -1 \leq t \leq 0$$

and, for $0 < t \leq 1$,

$$P(\min_{1\leq i\leq n} Y_i + \max_{1\leq i\leq n} Y_i \leq 2t) = 1 - P(\min_{1\leq i\leq n} Y_i + \max_{1\leq i\leq n} Y_i \leq -2t) = 1 - \frac{(1-t)^n}{2}.$$

Therefore, the density of T is given by

$$g(t) = \begin{cases} \frac{n}{2}(1+t)^{n-1} & \text{for } -1 \leq t \leq 0; \\ \frac{n}{2}(1-t)^{n-1} & \text{for } 0 < t \leq 1. \end{cases}$$

This shows that $g(t)$ is strictly increasing on $(-1,0)$, which proves that FR_g is strictly increasing on $(-1,0)$. Further,

$$\frac{g(t)}{1-G(t)} = \frac{n}{1-t} \qquad \text{for } 0 < t < 1,$$

which shows that FR_g is strictly increasing on $(0,1)$. The result then follows from the fact that FR_g is continuous on $(-1,1)$. $\square$

For the influence of the sample size on the risk function we need the notion of peakedness about zero of a random variable. From its definition in Section 4 it follows that the risk function of the rule δ_0 can be written as

$$R(\theta,\delta_0,\varepsilon) = \frac{1}{2}(\mu-\varepsilon)^p(1-p_{Z_2-Z_1}(\mu)), \tag{3.2}$$

where $p_{Z_2-Z_1}(\mu), \mu > 0$, is the peakedness of $Z_2 - Z_1$ about zero. (In what follows we will leave off the "about zero"). So, for a given choice of Z_1 and Z_2, the risk function of δ_0 is strictly decreasing in the common sample size n if the peakedness of $Z_2 - Z_1$ is strictly increasing in n.

The result of Birnbaum (1948) quoted in Lemma 4.1 reduces the behaviour of the peakedness $Z_2 - Z_1$ as a function of n to that of Z_1 and Z_2. More specifically we have

Lemma 3.4 *When the Z_i's have symmetric unimodal distributions, the peakedness of $Z_2 - Z_1$ is strictly increasing in n when the peakedness of each of Z_1 and Z_2 is strictly increasing in n.*

The question of when a summary statistic has increasing peakedness in n was, for the sample mean, answered by Proschan ((1965), Corollary 2.4). He proved the following result.

Lemma 3.5 *Let f be a PF_2 (Pólya frequency function of order 2) density, $f(y) = f(-y)$ for all y, $Y_1,\ldots,Y_n$ independently distributed with density f. Then $(1/n)\sum_{i=1}^n Y_i$ is strictly increasing in peakedness as n increases.*

The equivalences (4.4) in Section 4 tell us that a PF_2 density f is strictly unimodal and therefore unimodal. So, by the lemmas 3.4 and 3.5, in cases where the X_i are sample means, the risk function is for each $\mu \in (\varepsilon, a)$ strictly decreasing in n when f is PF_2. Or, equivalently, strictly unimodal, or logconcave on the interior of the support of F. Also, from (4.4), each of these properties implies that the $X_{i,j}$, and thus the sample mean, have IFR.

Nothing seems to be known about the behaviour of the peakedness of the median or the midrange as a function of n. We obtained the following two results.

Lemma 3.6 *Let $Y_1,\ldots,Y_n$ be independent and identically distributed with density f and let n be odd. Further, let M_n be the sample median and let $\mathcal{M} = [m_1, m_2]$ be the set of medians of F. Then, for x such that $\frac{1}{2} < F(m+x) < 1$, the peakedness of $M_n - m$ is, for $m \in \mathcal{M}$, strictly increasing in n.*

Proof. Assume without loss of generality that $m = 0$. First note that, for $x \in (-\infty, \infty)$,

$$P(M_n > x) = \sum_{i=0}^{(n-1)/2} \binom{n}{i} F(x)^i (1-F(x))^{n-i} = 1 - \frac{1}{B\left(\frac{n+1}{2}, \frac{n+1}{2}\right)} \int_0^{F(x)} t^{\frac{n-1}{2}} (1-t)^{\frac{n-1}{2}} dt.$$

So, as a function of $y = F(x)$, $0 < y < 1$,

$$\frac{d}{dy} P(M_n > x) = -\frac{y^{\frac{n-1}{2}}(1-y)^{\frac{n-1}{2}}}{B\left(\frac{n+1}{2}, \frac{n+1}{2}\right)}.$$

Putting $Q_n(y) = P(M_n > x) - P(M_{n+2} > x)$, this gives

$$\begin{aligned} \frac{d}{dy} Q_n(x) &= \frac{(n+2)!}{\left(\left(\frac{n+1}{2}\right)!\right)^2} y^{\frac{n+1}{2}} (1-y)^{\frac{n+1}{2}} - \frac{n!}{\left(\left(\frac{n-1}{2}\right)!\right)^2} y^{\frac{n-1}{2}} (1-y)^{\frac{n-1}{2}} \\ &= y^{\frac{n-1}{2}} (1-y)^{\frac{n-1}{2}} \frac{n!}{\left(\left(\frac{n+1}{2}\right)!\right)^2} \left((n+1)(n+2) y (1-y) - \left(\frac{n+1}{2}\right)^2 \right). \end{aligned}$$

This last expression is, for $0 < y < 1$, $> 0, = 0, < 0$ if and only if

$$G(y) = -y^2 + y - \frac{n+1}{4(n+2)} = \frac{1}{4(n+2)} - (y - \frac{1}{2})^2 \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} 0,$$

which is equivalent to

$$\left| y - \frac{1}{2} \right| \left\{ \begin{matrix} < \\ = \\ > \end{matrix} \right\} c = \frac{1}{2}\sqrt{(n+2)^{-1}}.$$

So, $Q_n(y)$ is increasing on $(\frac{1}{2} - c, \frac{1}{2} + c)$ and decreasing on $(0, \frac{1}{2} - c)$ and on $(\frac{1}{2} + c, 1)$. Combining this with the fact that, for all n,

$$P(M_n > x) = \begin{cases} 1 & \text{for } y = 0 \\ \frac{1}{2} & \text{for } y = \frac{1}{2} \\ 0 & \text{for } y = 1, \end{cases}$$

shows that

$$P(M_n > x) - P(M_{n+2} > x) \begin{cases} > 0 & \text{for } x \text{ such that } \frac{1}{2} < F(x) < 1 \\ < 0 & \text{for } x \text{ such that } 0 < F(x) < \frac{1}{2}, \end{cases}$$

which proves the result. □

For the midrange the following result holds.

Lemma 3.7 *For a sample $Y_1, \ldots, Y_n$ from a uniform distribution on the interval $[-1, 1]$, the peakedness of the midrange*

$$T = \frac{1}{2}(\min_{1 \le i \le n} Y_i + \max_{1 \le i \le n} Y_i)$$

is strictly increasing in n.

Proof. The result follows immediately from the proof of Lemma 3.3. □

Examples of cases where the density f is PF_2 are the normal, the double exponential, the uniform on the interval (a, b) and the logistic distribution. Given that these distributions are all symmetric and given the equivalences (4.4), Condition A(1) is satisfied when the sample means or the sample medians with n odd are used as summary statistics. This follows from the lemmas 3.1 and 3.2. It can easily be seen that Condition A(2) is also satisfied in these cases.

For the uniform distribution when using the midrange, Condition A(1) is satisfied by the lemmas 3.1 (part a)) and 3.3. That Condition A(2) is also satisfied is easily verified. Note that, for the case where the medians (with n odd) are used as summary statistics, Condition A(1) is satisfied when F is symmetric and has IFR. This follows from the lemmas 3.1 (part a)) and 3.2. As noted by Barlow, Marshall and Proschan (1963, p. 379), there do exist distributions F which have IFR but whose density f is not logconcave on the interior of the support of F. So, for this case, weaker conditions apply.

4 Logconcavity, Pólya frequency functions, strong unimodality, peakedness and IFR

In this section, definitions and results are assembled concerning the notions of logconcavity, Pólya frequency functions, total positivity, strong unimodality, peakedness and IFR.

We start with total positivity and Pólya frequency functions of order 2. These are defined as follows (see Schoenberg (1951)). Let $K(x, y)$ be defined on $A \times B$ where A and B are subsets of R. Then K is TP_2 (totally positive of order 2) if $K(x, y) \ge 0$ for all $x \in A$, $y \in B$ and, for all $x_1 \le x_2$, $y_1 \le y_2$, $x_i \in A$, $y_i \in B, i = 1, 2$,

$$\begin{vmatrix} K(x_1, y_1) & K(x_1, y_2) \\ K(x_2, y_1) & K(x_2, y_2) \end{vmatrix} \ge 0. \tag{4.3}$$

Schoenberg (1951) shows that $L(x - y) = K(x, y)$ is TP_2 if and only if

$$L(t) \ge 0 \text{ and } \log L(t) \text{ is logconcave on } R.$$

Here "logconcave on R" means, for a density f and corresponding distribution function F, that $\log f$ is concave on the interior of the support of F. A TP_2 function is also called a PF_2 (Pólya frequency function of order 2).
The notion of strong unimodality was introduced by Ibragimov (1956). He called a distribution function strongly unimodal if its convolution with every unimodal distribution function is unimodal, where a distribution function F is unimodal with mode m if $F(x)$ is convex for $x < m$ and concave for $x > m$. Ibragimov showed that a distribution function is strongly unimodal if and only if it is either degenerate or it is absolutely continuous with respect to Lebesgue measure and its density has a version which is logconcave (on the interior of the support of F).
Finally, if a density f is logconcave on the interior of the support of F, then the distribution function F has IFR. A proof of this can, e.g., be found in Marshall and Olkin (1979, p. 493).
Summarizing the above we get

$$
\begin{aligned}
& f \text{ is logconcave on the interior of the support of } F \Leftrightarrow \\
& f \text{ is strongly unimodal } \Leftrightarrow f \text{ is } \mathrm{PF}_2 \\
& \Rightarrow f \text{ has IFR.}
\end{aligned} \tag{4.4}
$$

The notion of "peakedness of a random variable W (about 0)" was introduced and studied by Birnbaum (1948). It is defined by

$$p_W(r) = P(|W| \leq r) \quad r > 0.$$

Further, W_1 is more peaked about 0 than W_2 if

$$p_{W_1}(r) > p_{W_2}(r) \quad \text{for all } r > 0.$$

Birnbaum (1948) proved the following result.

Lemma 4.1 *Let $W_i, 1 = 1, \ldots, 4$, be random variables with Lebesgue densities $f_i, i = 1, \ldots, 4$, respectively which are symmetric around 0 and such that*

i) For i=1, 3, W_i and W_{i+1} are independent;

ii) $f_2(w)$ and $f_3(w)$ are nondecreasing in w for $w > 0$;

iii) For $i = 1, 2$, W_i is more peaked about 0 than W_{i+2}.

Then $W_1 + W_2$ is more peaked about 0 than $W_3 + W_4$.

References

Barlow, R. E., Marshall, A. W. and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. Ann. Math. Statist., 34, 375-389.

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist., 25, 16-29.

Birnbaum, Z. W. (1948). On random variables with comparable peakedness. Ann. Math. Statist., 19, 76-81.

Ibragimov, I. A. (1956). On the composition of unimodal distributions. Theor. Probab. Appl., 1, 255-260.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications*, Academic Press.

Proschan, F. (1965). Peakedness of distributions of convex combinations. Ann. Math. Statist., 36, 1703-1706.

Schoenberg, I. J. (1951). On Pólya frequency functions I. J. Anal. Math., 1, 331-374.

Szekli, R. (1995). *Stochastic Ordering and Dependence in Applied Probability*, Springer-Verlag.

van der Laan, P. and van Eeden, C. (1998). On selecting the best of two normal populations using a loss function. Submitted.

Addresses for correspondence
Paul van der Laan
Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven, The Netherlands
e-mail: PvdLaan@win.tue.nl

Constance van Eeden
Moerland 19
1151 BH Broek in Waterland, The Netherlands
e-mail: cve@xs4all.nl