Article

Extreme Tail Ratios and Overrepresentation among Subpopulations with Normal Distributions

Theodore P. Hill 1,*† and Ronald F. Fox 2,†

1 School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA
2 School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA
* Correspondence: hill@math.gatech.edu
† These authors contributed equally to this work.

Abstract: Given several different populations, the relative proportions of each in the high (or low) end of the distribution of a given characteristic are often more important than the overall average values or standard deviations. In the case of two different normally-distributed random variables, as is shown here, one of the (right) tail ratios will not only eventually be greater than 1 from some point on, but will even become infinitely large. More generally, in every finite mixture of different normal distributions, there will always be exactly one of those distributions that is not only overrepresented in the right tail of the mixture but even completely overwhelms all other subpopulations in the rightmost tails. This property (and the analogous result for the left tails), although not unique to normal distributions, is not shared by other common continuous centrally symmetric unimodal distributions, such as Laplace, nor even by other bell-shaped distributions, such as Cauchy (Lorentz) distributions.

Keywords: tail ratio; normal distribution; overrepresentation; finite mixture of distributions

1. Introduction

In comparisons of the distributional values of a given psychological or physical trait between two populations (e.g., treated/untreated, male/female, exposed/nonexposed, elderly/youthful), the relative proportion of each population with values exceeding specified threshold or cutoff levels is often of more interest than comparisons of average values, standard deviations, or combinations thereof such as the Cohen \(d\) effect-size measure. The (right) tail ratio of one distribution compared to a second distribution, a measure of the relative tail proportions, is the ratio of the fraction of the first population above a given cutoff to the fraction of the second population above that same cutoff.

Tail ratios are a common measure of differences in extremes between populations in general and are of particular interest in psychological research, as emphasized by [1] in their review of the background, history, and practical advantages of tail ratios. For example, ref. [2] reported actual numerical ranges (between two and four) of certain male/female tail ratios and [3] studied how certain male/female tail ratios have changed over time. More generally, the issues of over- and underrepresentation of various factions of populations consisting of two or more subpopulations has become an important subject of study (see [4–8]).

The values of many biological or psychological traits (e.g., blood pressure, IQ, height) are often assumed to have normal (Gaussian) distributions (e.g., [9–11]), and the goal of this note is to record a simple fact about normal distributions that may be useful in interpreting statistical data concerning both tail ratios and over- and underrepresentation in mixed populations. In particular, it is shown below that in every population consisting of a finite number of subpopulations with different distributions of a normally distributed trait, exactly one of the subpopulations will not only dominate every other one in the right tail but also will do this in an extreme manner, eventually overwhelming all the other subpopulations. This property, although not unique to normal distributions, is not
shared by other common distributions, including ones that are also continuous, centrally symmetric, and unimodal, nor even by other bell-shaped distributions such as the common Cauchy (Lorentz) distributions.

2. Tail Ratios and Right-Tail Dominance

Recall that a probability measure $P$ on the real line is uniquely determined by its complementary cumulative distribution function (ccdf) $F_P$, defined by $F_P(x) = P((x, \infty))$ for all $x \in \mathbb{R}$. ($F_P$ is also often called the survival function of $P$, since $F_P(c)$ represents the $P$-probability of the set above the cutoff threshold $c$, i.e., the fraction that survives when all values less than or equal to $c$ are removed.) With this notation, the formal definition of the tail ratio of $P_1$ to $P_2$ is as follows.

**Definition 1.** Given probability distributions $P_1$ and $P_2$, the tail ratio of $P_1$ to $P_2$ is the function $R_{P_1}(c)/R_{P_2}(c)$, where $R_{P_i}$ and $F_{P_i}$ are the ccdfs for $P_i$ and $P_2$, respectively.

(N.B. By convention, the tail ratio of $P_1$ to $P_2$ is that of the right tail; i.e., the ratio of the ccdf of $P_1$ to the ccdf of $P_2$, not the ratio of the cdf of $P_1$ to the cdf of $P_2$, which would yield the left tail ratio.)

A distribution $P_1$ may be said to dominate another distribution $P_2$ in the right tail if, for all sufficiently large cutoffs $c$, the tail ratio of $P_1$ to $P_2$ is strictly greater than 1, i.e., $R_{P_1}(c) > R_{P_2}(c)$. A much stronger notion of domination in the right tail is when the survival probabilities of $P_1$ eventually become arbitrarily larger than those of $P_2$ as the cutoff increases; this is formalized in the next definition.

**Definition 2.** A probability distribution $P_1$ strongly dominates distribution $P_2$ in the right tail if the (right) tail ratio of $P_1$ to $P_2$ becomes infinitely large as the cutoff increases, i.e., if

$$\lim_{c \to \infty} \frac{R_{P_1}(c)}{R_{P_2}(c)} = \infty.$$

We also recall that a continuous (absolutely continuous) probability distribution $P$ is uniquely determined by its probability density function $f_P : \mathbb{R} \to [0, \infty)$ via $P((a, b)) = \int_a^b f_P(x) \, dx$; so, in particular, $R_{P}(c) = \int_c^{\infty} f_P(x) \, dx$. The next lemma, which records a simple relationship between the quotients of probability density functions (pdfs) and the quotients of the corresponding ccdfs, is used in several examples and proofs below.

**Lemma 1.** Suppose $P_1$ and $P_2$ are continuous probability distributions with strictly positive continuous pdfs $f_1$ and $f_2$ and with ccdfs $F_1$ and $F_2$, respectively. If $\lim_{x \to \infty} \frac{f_1(x)}{f_2(x)} = \alpha$, then $\lim_{c \to \infty} \frac{R_{F_1}(c)}{R_{F_2}(c)} = \alpha$.

**Proof.** Let $f_1$ and $f_2$ be strictly positive continuous pdfs with corresponding ccdfs $F_1$ and $F_2$, respectively. Then, since $\lim_{x \to \infty} \frac{F_1(x)}{F_2(x)} = 0$, and $f_2(x) > 0$ for all $x$,

$$\lim_{x \to \infty} \frac{F_1(x)}{F_2(x)} = \lim_{x \to \infty} \frac{F_1(x)}{F_2(x)} = \lim_{x \to \infty} \frac{f_1(x)}{f_2(x)} = \alpha,$$

where the first equality follows by the general form of L’Hôpital’s rule, the second since $F_1(x) = -f_1(x)$, and the third by hypothesis. □

The next example illustrates the difference between domination and strong domination.

**Example 1.** (i) Let $P_1$ and $P_2$ be Cauchy distributions with medians $m_1 = 0$ and $m_2 = 0.5$ and scale parameters $s_1 = 1$ and $s_2 = 0.5$, i.e., with density functions $f_{P_1}(x) = \left(\pi s_1^2 [1 + ((x - m_1)/s_1)^2]\right)^{-1}$, respectively. Then by Lemma 1,
The appropriateness of assuming that given data have a normal distribution is often tested using the well-known empirical observation called the “68%-95%-99.7% rule” of normality illustrated in Figure 1(left). The one key property of a continuous centrally symmetric unimodal distribution that makes it normal is the unique (after rescaling) rate of decrease in its density function away from its mean. The normal density function, discovered by Gauss in 1809 in connection with his studies of astronomical observation errors, decreases from its mean at a rate exactly proportional to $e^{-x^2}$ and not to $e^{-x}$ or $x^{-2}$, for example, as is the case for the Laplace and Cauchy distributions, respectively.
Figure 1. Universal empirical rules for all normal (left) and Cauchy (right) distributions. Here \( m \) represents the median, and \( s \) represents the distance (called standard deviation for normal distributions and scale parameter for Cauchy) from \( m \) to the inflection point. Although both families, normal and Cauchy, have similar bell-shaped density functions, normal distributions satisfy the main strong-dominance and overrepresentation properties presented in this note, but Cauchy distributions do not.

The Cauchy distribution, for instance, which sees widespread application in physics, also has a continuous centrally symmetric unimodal bell-shaped density similar to the normal distribution, but the Cauchy distribution has an undefined mean and variance and satisfies a different empirical rule, namely the 50%–70%–79.5% rule illustrated in Figure 1(right). Thus, it is very easy in practice to distinguish between these two similar-looking common bell-shaped distributions.

As is easy to see, the density functions of every two different normal distributions intersect in either exactly one or in exactly two distinct points. Thus, the density function of one of those two distributions is strictly larger than that of the other at all points greater than the larger of the two intersection points (or the unique one, if there is only one). This, in turn, implies that the proportion of that distribution from that point on is strictly larger than that of the other distribution from that point on; thus, this distribution will be overrepresented in the right tail. This is illustrated in the following example.

(N.B. For brevity, the standard notation \( N(m, \sigma^2) \) will be used throughout this note to denote a normal distribution with mean \( m \) and standard deviation \( \sigma > 0 \).)

**Example 2.**

(i) Let \( P_1 \sim N(100, 10^2) \) and \( P_2 \sim N(110, 10^2) \). It is clear that the unique crossing point of the density functions of \( P_1 \) and \( P_2 \) is at \( x = 105 \), which implies that the proportion of \( P_2 \) that is above any cutoff \( c > 105 \) is greater than the proportion of \( P_1 \) above \( c \), i.e., the tail ratio of \( P_2 \) to \( P_1 \) is greater than 1 for all cutoff values \( c \) strictly greater than 105. Conversely, the proportion of \( P_1 \) below any \( c < 105 \) is greater than the proportion of \( P_2 \) below \( c \).

(ii) Let \( P_1 \sim N(100, 10^2) \) and \( P_2 \sim N(101, 11^2) \). By basic algebra, the two crossing points of the density functions of \( P_1 \) and \( P_2 \) are seen to be at \( x_1 \approx 83.52 \) and \( x_2 \approx 106.95 \), which implies that the tail ratio of \( P_2 \) to \( P_1 \) is greater than 1 for all cutoffs \( c > x_2 \). Similarly, in this case \( P_2 \) also dominates \( P_1 \) in the lower tail in that the proportion of \( P_2 \) that is below any cutoff \( c < x_1 \) is also greater than the proportion of \( P_1 \) below \( c \).

As was seen in Example 1, for two given different Cauchy distributions or two different Laplace distributions, neither distribution may strongly dominate the other in the right tail. This is in sharp contrast to the main conclusion in this note, where it will be shown that in every finite collection of different normal distributions, there is always a unique one of those distributions that strongly dominates every one of the other distributions in the right tail.
Theorem 1. Let $P_1$ and $P_2$ be different normal distributions. Then,

(i) either $P_1$ strongly dominates $P_2$ in the right tail or $P_2$ strongly dominates $P_1$ in the right tail;
(ii) if $P_1$ strongly dominates $P_2$ in the right tail, then either $P_1$ has greater mean (average value) than $P_2$ or $P_1$ has greater variance than $P_2$ or both;
(iii) if $P_1$ has greater variance than $P_2$, then $P_1$ strongly dominates $P_2$ in both right and left tails, independent of the means.

Proof. Suppose that $P_1 \sim N(m_1, \sigma_1^2)$ and $P_2 \sim N(m_2, \sigma_2^2)$ are normal distributions with pdfs $f_1$ and $f_2$, respectively. Since $P_1$ and $P_2$ are different, either $\sigma_1 \neq \sigma_2$ or $\sigma_1 = \sigma_2$ and $m_1 \neq m_2$.

Case 1. $\sigma_1 \neq \sigma_2$. Without loss of generality, $\sigma_1 > \sigma_2 = 1$. Then,

$$
\frac{f_1(x)}{f_2(x)} = \frac{\sqrt{2\pi}}{\sigma_1 \sqrt{2\pi}} \exp \left( \frac{-(x - m_1)^2}{2\sigma_1^2} + \frac{(x - m_2)^2}{2} \right) = \frac{1}{\sigma_1} \exp \left( \frac{x^2}{2} \left( 1 - \frac{1}{\sigma_1^2} \right) + \frac{m_1^2 - m_2^2}{\sigma_1^2} + \frac{m_1^2 - m_2^2}{2\sigma_1^2} \right) .
$$

Since $\frac{1}{\sigma_1^2} > 1$, $\left( 1 - \frac{1}{\sigma_1^2} \right) > 0$, then (1) implies that $\lim_{x \to \infty} (f_1(x) / f_2(x)) = \infty$, and thus by Lemma 1, $P_1$ strongly dominates $P_2$ in the right tail.

Case 2. $\sigma_1^2 = \sigma_2^2$ and $m_1 \neq m_2$. Without loss of generality, $\sigma_1 = \sigma_2 = 1$. Then,

$$
\frac{f_1(x)}{f_2(x)} = \exp \left( \frac{-x^2 + 2m_1x - m_1^2 + x^2 - 2m_2x + m_2^2}{2} \right) = \exp \left( (m_1 - m_2)x + \frac{(m_2^2 - m_1^2)}{2} \right) .
$$

In addition, without loss of generality, $m_1 > m_2$, in which case (2) implies, via Lemma 1 as in Case 1, that $P_1$ strongly dominates $P_2$ in the right tail. This concludes the proof of (i); the proofs of (ii) and (iii) follow similarly.

The same essential argument extends easily to show that among every finite collection of different normal distributions, strong domination in the right tail by exactly one of those distributions is inevitable.

Corollary 1. Given a finite number of different normal distributions $P_1, \ldots, P_n$, there is a unique one of these distributions that strongly dominates all the others in the right tail.

Proof. For each $i \in \{1, \ldots, n\}$, let normal distributions $P_i$ have mean $m_i$ and standard deviation $\sigma_i$, respectively. Since the distributions are all different, if $m_i = m_j$ and $\sigma_i = \sigma_j$ then $i = j$, which implies that there exists a unique $i^* \in \{1, \ldots, n\}$ such that $m_{i^*} = \max \{m_1, \ldots, m_n\}$: $\sigma_{i^*} = \max \{\sigma_1, \ldots, \sigma_n\}$. By the arguments for Cases 1 and 2 in Theorem 1, $P_{i^*}$ strongly dominates $P_i$ in the right tail for all $i \neq i^*$.

As was seen in Theorem 1, if $P_1$ has either greater variance than $P_2$ or the same variance and higher mean, then $P_1$ will strongly dominate $P_2$ in the right tail. Moreover, for many practical purposes, “most important, what might appear to be trivial group differences in both variability and central tendency can cumulate to yield very appreciable differences between the groups in numbers of extreme scorers” ([10], p. 11). The next example, a slight modification of the numerical example suggested by Feingold, illustrates this observation with the two normal distributions in Example 2(ii) that are close in mean value (100 vs. 101) and in standard deviation (10 vs. 11).

Example 3. Suppose a population $X$ consists of two mutually exclusive subpopulations $X_1$ and $X_2$, where the values of a given trait are normally distributed with distributions $P_1 \sim N(100, 10^2)$ and $P_2 \sim N(101, 10^2)$.
and $P_2 \sim N(101, 112)$, respectively, as in Example 2(ii). A normal distribution calculator yields that $P(X_1 > 130) \approx 1.3490 \times 10^{-3}$ and $P(X_2 > 130) \approx 4.1900 \times 10^{-3}$, which yields the tail ratio $P(X_2 > 130)/P(X_1 > 130) \approx 3.106$. Thus, even with distributions this close in average value and standard deviation, if the two subpopulations $X_1$ and $X_2$ are of the same size, then the $X_2$ population will comprise more than 75% of the combined population beginning only three standard deviations above the mean.

Note that the results in Theorem 1 and Corollary 1 for normal distributions only depend on first-order asymptotic terms, and the question of which more general classes of distributions with rapidly decaying tails satisfy the same conclusions is left to the interested reader.

### 4. Overrepresentation in the Right Tail

Whether a particular subpopulation is overrepresented or underrepresented among the other subpopulations with respect to given values for a specific trait depends on the relative size of that subpopulation with those trait values compared to the size of the whole population with those trait values. For example, if subpopulation $X_1$ comprises 30% of the total population, but comprises 40% of the population with trait values above a given cutoff $c$, then $X_1$ is overrepresented in the portion of the total population with values greater than $c$.

The goal of this section is to show that a simple consequence of Corollary 1 is that in every finite mixture of different normal distributions, exactly one of those distributions will be strongly overrepresented in the right tail. (Recall that a finite mixture of distributions is a probability distribution with cdf $F = \sum_{i=1}^{n} w_i F_i$, where $n > 1$, $F_1, \ldots, F_n$ are cdfs, and $w_1, \ldots, w_n$ are strictly positive weights with $\sum_{i=1}^{n} w_i = 1$.)

**Definition 3.** Given a finite mixture of distributions $F = \sum_{i=1}^{n} w_i F_i$, the distribution $F_i$ is strongly overrepresented in the right tail of $F$ if, as $c \to \infty$, the proportion of subpopulation $F_i$ with values above $c$ approaches 100% of the total population of $F$ with values above $c$, that is, if

$$\lim_{c \to \infty} \frac{w_i F_i(c)}{\sum_{i=1}^{n} w_i F_i(c)} = 1.$$ 

**Theorem 2.** In every finite mixture of different normal distributions $F = \sum_{i=1}^{n} w_i F_i$, there is a unique $i^* \in \{1, \ldots, n\}$ such that $F_{i^*}$ is strongly overrepresented in the right tail of $F$.

**Proof.** For all $i \in \{1, \ldots, n\}$, let $F_i$ denote the ccdf of $P_i$. By Corollary 1, there exists a unique $i^* \in \{1, \ldots, n\}$ such that $P_{i^*}$ strongly dominates $P_i$ for all $i \neq i^*, i \in \{1, \ldots, n\}$, i.e.,

$$\lim_{c \to \infty} \frac{F_i(c)}{F_{i^*}(c)} = 0 \quad \text{for all } i \neq i^*, i \in \{1, \ldots, n\}. \quad (3)$$

Since $w_i > 0$ for all $i \in \{1, \ldots, n\}$, (3) implies that

$$\lim_{c \to \infty} \frac{w_i F_i(c)}{w_{i^*} F_{i^*}(c)} = 0 \quad \text{for all } i \neq i^*, i \in \{1, \ldots, n\},$$

which implies that

$$\lim_{c \to \infty} \frac{\sum_{i=1}^{n} w_i F_i(c)}{\sum_{i=1}^{n} w_i F_i(c)} = 1.$$
so $F_{so}$ is strongly overrepresented in the right tail of $F$. □

Two concrete examples of normally-distributed traits among human populations are height and test scores such as those in the College Board Scholastic Aptitude Test (SAT). There are enormous amounts of data on human height, which are essentially continuous and are very close to being normally distributed ([11], p. 24). Scores on the SAT, on the other hand, are originally discrete but the distribution “obtained from a continuized, smoothed frequency distribution of original SAT scores” is a linear transformation of a normal distribution ([9], p. 59). Thus, since all linear transformations of normal distributions are normal, for all practical purposes, the resulting smoothed SAT scores have normal distributions.

Example 4. The SAT scores of males and females are usually assumed (or designed) to be approximately normally distributed [9]. Unless the distributions are identical, Theorem 2 implies that exactly one of those two sexes must be strongly overrepresented in the right tail, and that this overrepresentation will increase as the score range increases; Figure 2 illustrates this with actual College Board statistics.

![Table](image)

**Figure 2.** Figure from [13], p. 7, titled “The numbers College Board didn’t publish” showing statistics for nearly two million students for the 2016 Edition of the Scholastic Aptitude Test, with breakdown by gender and score ranges. Note that the proportions of males in various score ranges, i.e., the tail ratios, increase as the score range increases and the left tail ratios also increase as the score range decreases. (About 10% more females participated than males, which is reflected in the Adjusted Male/Female Ratios).

5. Discussion

In real life examples, of course, there are no variables that are exactly normally distributed, since the normal distribution is continuous, and the number of people in various categories, for example, is necessarily finite. However, if distributions are close to being normally distributed, the right-tail overrepresentation of a unique subpopulation predicted by Theorem 2 (and the analogous conclusions for left tails) are perhaps reasonable to expect. Similarly, in real-life examples, calculations involving tails that are 6 or 7 standard deviations out involve probabilities of less than one in 10 billion and are meaningless among the current human population of this planet.

Author Contributions: Both authors contributed equally to this work and have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.
References


3. Lakin, J.M. Sex differences in reasoning abilities: Surprising evidence that male-female ratios in the tails of the quantitative reasoning distribution have increased. *Intelligence* **2013**, *41*, 263–274. [CrossRef]


10. Feingold, A. The additive effects of differences in central tendency and variability are important in comparisons between groups. *Am. Psychol.* **1995**, *50*, 5–13. [CrossRef]

