A Follow-up to the Range Disparity Distribution and Its Applications

Lawrence N. Marinucci[1], Paul Lupinacci[2], Joel Waksman[3], Tai Xie[3]

[1]Acorda Therapeutics, Inc., 420 Saw Mill River Road, Ardsley, NY 10502
[2]Villanova University, 800 E. Lancaster Avenue, Villanova, PA 19085
[3]Brightech International, 285 Davidson Avenue, Somerset, NJ 08873

Abstract
Conventionally, inferential statistics focuses on measures of central tendency, eg, the population mean. Little attention is given to differences in the range of sample distributions, and the range is often associated with exceptional values. Under certain circumstances, effects visible at the edges of a sample distribution may provide more sensitive and even more meaningful measures of differences between populations than the difference in means. In this paper we introduce the Range Disparity Distribution (RDD) to describe the probability that a given number of items in one set fall outside the range of all the items in another set on a given measure. Here we define the RDD, prove that it is a probability distribution, and then proceed to derive the mean and variance. We also provide a statistical test comparing two groups based on the RDD.

Key Words: probability distribution, central tendency, range, sample distributions, exceptional values

Introduction

At the 2007 American Statistical Association meetings, "The Range Disparity Distribution and Its Applications" was presented as a contribution to "Application of Innovative Analysis in Clinical Trials." In that presentation, we introduced an analysis of range disparity [1]. Our position was that for the purpose of comparing groups—for example, groups of patients undergoing differing treatments in clinical trials—little attention had been devoted to the ranges of clinical data. This led us to formulate a statistical test for the number of values in one dataset expected to be beyond the range of all the values in another, independent dataset. In certain circumstances, such an analysis may provide more sensitive and even more meaningful measures of differences between populations than the differences in means. Several practical examples were given. This paper notes a special-case relation between the range disparity distribution (RDD) and the negative hypergeometric distribution [2–4], and from that derives the mean and variance of the RDD. As further practical examples, we describe potential uses of the RDD in analyses of clinical and industrial data.

Illustrating the Basic Problem

Suppose that of 10 straws with randomly differing lengths, 6 happen to be in urn *A* and the other 4 in urn *B*. What is the probability that a given number of the straws in urn *B* are longer than the longest straw in urn *A*? For urn *B* to contain no such straws, the answer requires no calculation. The 6 straws in urn *A* would have to include the longest straw of all 10, and for this to be the case, the probability is simply 6/10, or 60%. For urn *B* to contain exactly *one* straw longer than the longest straw in urn *A*,

- Urn *B* would have to include the longest straw, a 4-in-10 chance, and *also*:
- One of the 6 straws in urn *A* would have to be the longest among the remaining 9, a 6-in-9 chance.

The probability of both these requirements being met is 4/10×6/9, or 27%. For urn *B* to contain exactly *two* straws longer than the longest straw in urn *A*,

- Urn *B* would have to include the longest straw, a 4-in-10 chance, and *also*:
- Urn *B* would have to include the longest straw among the remaining 9, a 3-in-9 chance, and *also*:
- One of the straws in urn *A* would have to be the longest among the remaining 8, a 6-in-8 chance.

The probability of all these requirements being met is 4/10×3/9×6/8, or 10%. The full use of this logic produces the probability distribution displayed in **Table 1**.

Table 1: Probability Distribution for Likelihood of Longer Straws in a Set of 4 Than the Longest Straw in a Set of 6[a]

Number of Longer Straws	Probability Calculation	Resulting P Value
0	6/10	60%
1	4/10×6/9	27%
2	4/10×3/9×6/8	10%
3	4/10×3/9×2/8×6/7	3%
4	4/10×3/9×2/8×1/7×6/6	0.5%

[a]Due to rounding, the sum of all probabilities does not add up to exactly 100%.

Methods

The combinatorics of straws in urns suggests the following formal definition:

Let *S* and *T* be positive integers. Let *Y* be a random variable, and let *y* be an integer ranging from 0 to *T*. *P(Y=y)* has the distribution:

$$P(Y=y) = \begin{cases} \dfrac{S}{S+T}, & y=0 \\[2ex] \dfrac{S}{S+T} \times \displaystyle\prod_{k=1}^{y} \dfrac{T-k+1}{S+T-k}, & 0<y\leq T \end{cases}$$

$P(Y=y)$ represents the probability that a given number of items in a set of T values fall outside the maximum (or minimum) value in an independent set of S values. In combinatorial notation, $P(Y=y)$ can be rewritten as:

$$P(Y=y) = \begin{cases} \dfrac{C_{T-y}^{S+T-(y+1)}}{C_T^{S+T}}, & y=0, 1, ..., T \end{cases}$$

The negative hypergeometric distribution contains 3 parameters, S and T as above and $r=1, ..., S$, with Y as the random variable, as above. We have:

$$P(Y=y) = \binom{y+r-1}{r-1}\binom{S+T-y-r}{S-r} \Big/ \binom{S+T}{S}, \quad y=0, 1, ..., T$$

By setting r to 1, the negative hypergeometric distribution reduces to the RDD. From the mean and variance of the negative hypergeometric distribution, the mean and variance of the RDD are:

$$E(Y) = \frac{T}{S+1}$$

$$Var(Y) = \frac{TS(S+T+1)}{(S+1)^2(S+2)}$$

Statistical Inference Based on RDD

Having considered the basic properties of the RDD, we can examine application to the direct comparison of two groups. The capability of comparing groups at the upper and lower limits is quite important for evaluation of complex systems that are typical of biological or clinical testing, as some variables may only, or at least disproportionately, affect the upper tail of a distribution, the determining factors for the lower range residing elsewhere.

Suppose we have two independent groups and wish to compare the maximum values between the two groups. Let $Y_1, Y_2, ..., Y_T$ be the measures from Group 1 and $X_1, X_2, ..., X_S$ be the measures from Group 2. Assume that the measures are from a continuous distribution such that ties are not possible. Let $X_{max} = \max\{X_j: j=1, 2, ...,S\}$. Then $G = \sum_{i=1}^{T} I\{Y_i > X_{max}\}$ is the number of observations in Group 1 that are larger than the largest observation in Group 2. Thus, G has the RDD with:

$$E(G) = \frac{T}{S+1} \quad \text{and} \quad Var(G) = \frac{TS(S+T+1))}{(S+1)^2(S+2)}$$

A hypothesis can be formed as:

Null hypothesis: $H_O: \theta_1 = \theta_2$

Alt hypothesis: $H_A: \theta_1 > \theta_2$

Test statistic: $$G = \sum_{i=1}^{T} I\{Y_i > X_{max}\}$$

Decision rule: Reject H_O if $G \geq G_\alpha$ where G_α is chosen to satisfy $\Pr(G \geq G_\alpha) \leq \alpha$

where θ_1 and θ_2 represent the maximum values for the two populations. We can also form statistical inferences on group differences in terms of lower bounds by using the minimum and reversing the relationship in the null and alternative hypotheses.

"Subjects" and "Procedures"

Example 1.

Now consider a hypothetical clinical trial in which 20 subjects are randomized to take an active drug or placebo (10 subjects per group). Because hypercholesterolemia is a potential active-drug side effect, all subjects are required to have a baseline serum cholesterol level of 85 to 200 mg/dL.

Results

Table 2 lists the hypothetical subjects' post-baseline cholesterol levels. To facilitate comparison between the active-drug and the placebo group, the listings (and subject ID numbers) are by ascending cholesterol level.

Table 2: Post-baseline Cholesterol Levels in a Hypothetical Clinical Trial

Active-Treatment Group		Placebo Group	
Subject ID	Cholesterol (mg/dL)	Subject ID	Cholesterol (mg/dL)
001	125	002	126
003	138	005	138
004	166	007	168
006	178	008	172
009	190	010	188
011	192	013	192
012	203	014	195
015	214	016	198
017	223	018	201
020	231	019	202

Table 3 includes 3 standard statistical analyses of the post-baseline data. By all 3 methods, there is no basis for concluding that the active treatment raises cholesterol levels.

These results (**Table 3**) illustrate how standard analyses of lab data can miss potential signals. The conclusion from the data presented above would suggest that there is insufficient evidence to reject the null hypotheses of no treatment difference in favor of the alternative that the active treatment raises cholesterol; one might conclude that there is no drug safety concern.

In contrast to the standard methods, the RDD-based approach (also included in **Table 3**) begins by noting that the maximum cholesterol level in the placebo group was 202 mg/dL, and that 4 of the subjects in the active-treatment group had levels exceeding this value. By RDD calculation, the probability of finding that 4 or more values in a group of 10 exceed the maximum value in another group of 10 is 0.043. Thus, the RDD analysis may lead to the conclusion that risk for hypercholesterolemia is of clinical concern.

Table 3: Standard and RDD-Based Analyses of the Post-baseline Data

Variable	Active-Treatment Group (N=10)	Placebo Group (N=10)	Analysis Method	P Value[a]
Mean (SD) serum cholesterol, mg/dL	186.0 (35.0)	178.0 (26.9)	t-test	0.287
			Wilcoxon rank-sum test	0.289
n (%) with levels >200 mg/dL	4 (40%)	2 (20%)	Fisher's exact test	0.314
n (%) of active-treatment group with levels exceeding the range in the placebo group	4 (40%)	NA	RDD	0.043

[a]One-sided.
NA, not applicable; RDD, range disparity distribution; SD, standard deviation.

An important difference between standard methods and the RDD-based approach is that the standard methods often require a pre-selected threshold value for clinical meaningfulness (eg, 200 mg/dL, for use of Fisher's exact test in **Table 2**). In the RDD-based approach, the data themselves set a threshold.

Interestingly, if the threshold set is > 202 mg/dL (any value higher than the maximum value in the placebo group), the resulting p-value from Fisher's exact test comparing 4/10 active to 0/10 placebo is 0.043 (ie, the same p-value from the RDD; further analysis on the nature of this equality is needed). This says something very interesting about selecting somewhat arbitrary thresholds versus the data indicating what that threshold may be.

Example 2.

In cancer trials, the tumor response is defined as at least 25% reduction from baseline in the sum of the bi-dimensional longest diameters for target lesions ie, the "tumor burden" [5]. Tumor response is commonly used as a primary endpoint to obtain accelerated regulatory approval. In a refractory study, the patient has previously failed to respond to the standard therapy (ie, at this point, it would be assumed impossible for a 100% reduction in the standard therapy arm). Frequently, there are no overall group differences in the average response, but there is an obvious advantage in the tail of the distribution of the tumor burden. For example, there may be a number of subjects in the treatment arm who had tumor burden reductions greater than the largest reduction in the standard therapy arm.

Nonmedical Applications

Example 3.

In nonmedical as well as medical settings, 2 sets of data may have no significant difference in average value, but the tails of the distributions may disclose a pronounced difference. Suppose an industrial corporation suspects that its plant *A* is manufacturing items with a breaking strength lower than that of items manufactured by plant *B*. A limited sample consisting of 5 items from each of the plants is available for testing. If the breaking strength of more than 2 of the items from plant *A* falls below the range for all 5 items from plant *B*, there is a basis for suspicion, in that by RDD calculation, the likelihood of such a finding resulting only from chance variation would be less than 0.10. If 4 or 5 of the samples from plant *A* were below the failure range for the samples from plant *B*, the likelihood that the finding resulted from chance variation would be less than 0.025.

Conclusions

The RDD describes the probability that for a given measure, a given number of items in one set falls outside the range of all the items in another set. In this way, the RDD can test for consistency between the sets without a need to pre-establish an arbitrary threshold value for distinguishing between significant difference and chance oscillation. The RDD also offers a capacity for independent testing of the upper and lower bound of a dataset.

Acknowledgments

Conflict of Interest

References

[1] Marinucci LN, Xie T, Cohen R, Blight AR. The range disparity distribution and its applications. Presented at the American Statistical Association Joint Statistical Meeting [Biopharmaceutical Section]; July 29–August 2, 2007; Salt Lake City, UT. Available at: http://www.amstat.org/meetings/jsm/2007/onlineprogram/index.cfm?fuseaction=abstract_details&abstractid=309056 (accessed 20 September 2013).

[2] Wilks SS. Mathematical statistics. New York: John Wiley and Sons, Inc.; 1962, pp. 141–3.

[3] Johnson LJ, Kemp AW, Kotz S. Univariate discrete distributions. 3rd ed. Hoboken, NJ: John Wiley and Sons, Inc.; 2005, pp. 252–6.

[4] Miller GK, Fridell SL. A forgotten discrete distribution? Reviving the negative hypergeometric model. Am Stat 2007;61:347–50.

[5] Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 2000;92:205–16.