

QUANTIFYING DIFFERENCES IN PARASITE NUMBERS BETWEEN SAMPLES OF HOSTS

Bernhard Klar, Trevor N. Petney*, and Horst Taraschewski*

Institute for Stochastics, Karlsruhe Institute of Technology (KIT), Kaiserstrasse 89, 76133 Karlsruhe, Germany. e-mail: Bernhard.Klar@kit.edu

ABSTRACT: An important question in many parasitological studies is the assessment of differences in parasite numbers between samples of hosts. This is not always an easy problem with which to deal. While almost everyone will agree that the main task consists in deciding whether the values in one sample tend to be higher than the values of another sample, there is considerable disagreement about what higher (or lower) should mean. In most cases, dissimilarity measures are differences between mean values, medians, geometric means, prevalence rates, relative effects, and more. In general, different measures can lead to different conclusions. However, a debate as to which measure is superior is fruitless; it depends on goals and circumstances of the respective study. In our opinion, it is more important to identify situations in which most of the above measures coincide, and, hence, one can confidently claim that the values in one sample are higher than in another. This is the case when one sample is stochastically larger than a second. Roughly speaking, a random variable X is stochastically larger than a second random variable Y , when X assumes large values with a higher probability than Y . In this paper, we review this concept using distributional and data examples, and we propose the use of graphical tools for detecting stochastic dominance. Our method provides a relatively easy, visual way of fully justifying general statements that the number of parasites in one sample is larger than in another.

In much parasitological research, the quantification and comparison of parasite numbers in different samples, recorded at different times and/or in different habitats, is an important task. The most commonly used measure for this purpose is probably the difference between mean values, i.e., mean abundance or mean intensity. Rózsa et al. (2000) argue that terminological recommendations require the use of these measures. Because small sample sizes can lead to difficulties when using the traditional t -test or Welch's modification, they propose a bootstrap test based on Welch's statistic. They criticize the use of nonparametric tests such as the Mann-Whitney U -test because these tests compare other characteristics instead of means. Indeed, assuming we have samples X_1, \dots, X_m and Y_1, \dots, Y_n of size m and n , the U -test corrected for ties is based on the statistic

$$U = \sum_{i=1}^m \sum_{j=1}^n \begin{cases} 1 & \text{if } X_i > Y_j \\ \frac{1}{2} & \text{if } X_i = Y_j \end{cases}$$

(see, e.g., Hollander and Wolfe, 1999). $U/(m \cdot n)$ is an estimator of the relative effect $p = \Pr(X > Y) + (1/2) \Pr(X = Y)$. If $p > 1/2$, observations tend to be larger in group 1 in comparison to group 2 in a certain sense, but the mean of group 1 is not necessarily larger than the mean of group 2. The converse is true for $p < 1/2$. If both groups have the same distribution, then $p = 1/2$. However, the opposite does not hold: if $p = 1/2$, the underlying distributions need not be the same.

In 2 recent articles, however, Neuhäuser and Poulin (2004) and Neuhäuser and Ruxton (2009) advocate the use of relative effect, because "means are not very useful descriptors for skewed distributions," (Neuhäuser and Poulin, 2004) and they propose the use of the Brunner and Munzel (2000) test, which is also based on this measure. Neuhäuser and Poulin (2004) conclude that the relative effect would be the natural measure for a difference between 2 samples because "the main question is whether the values in one sample tend to be larger (or smaller) than the values of the other sample," implying that only the relative effect provides the answer to this question. A bootstrap test for the

hypothesis $p = 1/2$ based on the rank Welch test statistic is proposed by Reiczigel et al. (2005). A modified version of the Brunner-Munzel test is discussed in Neubert and Brunner (2007).

There are other measures in common use, such as the comparison of medians or prevalence rates, and, clearly, in a specific situation, the use of different measures can lead to different conclusions. However, in our opinion, a dispute about the "best" measure is fruitless because this depends on goals and circumstances of the respective study. For example, if an infection with a certain parasite is invariably lethal, prevalences are adequate measures. In other cases, infection with a moderate number of parasites has no demonstrable negative effects on the host, but a very large number of parasites cause significant morbidity; in such cases, means that give more weight at high intensities are suitable measures. Situations that lie in between call for intermediate measures such as medians or relative effects.

Unfortunately, because the impact of parasites on hosts is seldom known precisely, it is difficult to justify the preference for a particular measure, even in specific situations. For this reason, it is important to identify situations in which the above mentioned measures coincide; one can, therefore, confidently claim that the values in one sample are higher than in another. Thus, it is not sufficient to compare a special characteristic of 2 distributions, one must compare the whole distributions. A suitable tool is to determine whether the 2 distributions are stochastically ordered. We review this concept using distributional and data examples, and we propose graphical tools for detecting stochastic dominance.

STOCHASTIC DOMINANCE BETWEEN DISCRETE DISTRIBUTIONS

The (usual) stochastic order is an established concept in probability theory and is in frequent use in reliability theory and econometrics. Because we are interested in comparing parasite numbers, we will give the definition for discrete distributions: A discrete random variable X with cumulative distribution function (cdf) $F_X(k) = \Pr(X \leq k)$ is said to be larger than another discrete random variable Y with cdf $F_Y(k) = \Pr(Y \leq k)$ with respect to *stochastic order* (written $X \geq_{st} Y$) if

$$P(X > k) \geq P(Y > k) \quad \text{for } k = 0, 1, 2, \dots, \tag{*}$$

TABLE I. Differences in (cumulative) relative frequencies for example 1.

k	0	1	2	3	4	5	6	7	8	>8
Abs. freq. sample 1	32	15	8	4	1	1	3	2	1	4
Abs. freq. sample 2	134	19	9	0	1	2	0	1	1	1
$r_Y(k) - r_X(k)$	0.35	-0.1	-0.06	-0.06	-0.01	0	-0.04	-0.02	-0.01	-0.05
$R_Y(k) - R_X(k)$	0.35	0.25	0.19	0.13	0.12	0.12	0.08	0.06	0.05	0

or equivalently

$$1 - F_X(k) \geq 1 - F_Y(k) \quad \text{for } k = 0, 1, 2, \dots,$$

and it is called *strictly larger* if \geq can be replaced by $>$ for at least 1 k (otherwise, the 2 distributions can be equal). This means that if X stochastically dominates Y , then X assumes large values with a higher probability than Y does, and hence X assumes small values with a smaller probability than Y does. We also say that X is stochastically larger than Y , that X stochastically dominates Y , or that Y is stochastically smaller than X . Obviously, a further equivalent condition for $X \geq_{st} Y$ is given by $F_Y(k) - F_X(k) \geq 0$ for all k .

For our purpose, the most important properties of this order between random variables are the following:

- If $X \geq_{st} Y$, then X assumes values larger than 0 with a higher probability than Y does. Hence, the (theoretical) prevalence $P(X \geq 1)$ of X is larger than or equal to the prevalence of Y . Formally, this follows directly from the definition in (*) putting $k = 0$: $P(X \geq 1) = P(X > 0) \geq P(Y > 0) = P(Y \geq 1)$.
- If $X \geq_{st} Y$, then $EX \geq EY$. This follows from the definition of stochastic order together with the representation of the population mean $EX = \sum_{k \geq 0} k \cdot P(X=k) = \sum_{k \geq 0} P(X > k)$. If X stochastically dominates Y , then $P(X > k)$ is greater than or equal to $P(Y > k)$ for each k , and, hence, $EX \geq EY$.
- If $X \geq_{st} Y$, then all population quantiles of the distribution of X are larger than or equal to the corresponding quantiles of Y . In particular, the median of X is larger than or equal to the median of Y .
- If $X \geq_{st} Y$, then $p \geq \frac{1}{2}$ holds for the relative effect.

Hence, if $X \geq_{st} Y$, the typical measures for assessing differences between samples all point in the same direction. Examples for discrete distributions that are stochastically ordered will be dealt with below.

If the implications of stochastic order for empirical data are considered, and if we examine data sets, probabilities $P(X = k)$ are replaced by relative frequencies $r(k)$, $k = 0, 1, \dots$, and the cdf is replaced by the cumulative relative frequencies, which are sometimes called the empirical distribution function (edf). Because the edf has all the properties of a cdf, the 4 properties mentioned above carry over immediately to the corresponding empirical measures. For example, if $R_Y(k) - R_X(k) \geq 0$ for all k , (sample) prevalence and sample mean of the x-sample are larger than or equal to the corresponding measures of the y-sample.

The first example compares counts of larvae of swimbladder nematodes in 2 populations of the Japanese eel (*Anguilla japonica*) from southwest Taiwan. Münderle et al. (2006) compared wild eels from the Kao-Ping River (sample 2, $n = 168$) to cultured eels from an adjacent aquaculture farm (sample 1, $n = 71$). All

recorded nematodes were species of *Anguillicoloides crassus* (previously *Anguillicola crassus*).

Absolute frequencies and differences in relative frequencies are given in Table I. Figure 1 shows histograms of relative frequencies truncated at 22. We see that $r_X(0) < r_Y(0)$, but $r_X(1) > r_Y(1)$, $r_X(2) > r_Y(2)$, etc. However, it is not possible to deduce anything regarding stochastic ordering between the 2 samples based directly on these values.

The differences in cumulative relative frequencies are given in the last row of Table I. Because all entries are positive, sample 1 is stochastically larger than sample 2. This can also be seen by careful comparison of the 2 histograms of cumulative relative frequencies in Figure 2. From the above discussion, we can conclude that the difference between prevalences, means, and medians are positive and that the relative effect is larger than 1/2. The empirical values are given in the second column of Table II. In this example, it is fully justified to state that the eels from the aquaculture farm are more heavily infected than the wild eels.

GRAPHIC TOOLS FOR DETECTING STOCHASTIC ORDER

Instead of comparing 2 histograms of cumulative relative frequencies (or cumulative probabilities) as done in the first example, it is more helpful to visualize both quantities in a single graph. From the definition, 2 simple graphical procedures are evident for checking the stochastic order between 2 given discrete distributions. The first is a bar plot of the difference $F_Y(k) - F_X(k)$ with positive bars if X stochastically dominates Y , and negative values if Y stochastically dominates X . This plot is called a probability difference plot (P-D plot) in what follows. Alternatively, a scatter plot of $F_Y(k)$ against $F_X(k)$ can be drawn. If one plots the values of $F_X(k)$ on the x-axis, the points are above the diagonal if X stochastically dominates Y . This plot is known as probability-probability plot (P-P plot).

A further possible diagnostic plot is a quantile-quantile plot, which is also well known in the graphical analysis of statistical data. However, for discrete distributions, it is less suitable, particularly due to the need for choosing suitable plotting positions.

In all plots, $F_Y(k)$ and $F_X(k)$ must be replaced with $R_Y(k)$ and $R_X(k)$ when dealing with observed data instead of probability distributions.

The P-D plot and the P-P plot for the eel nematode example 1 are shown in Figures 3 and 4, respectively. As expected from the discussion regarding the eel nematodes, all values in Figure 3 are positive, and all points in Figure 4 are well above the diagonal. Hence, each plot shows at a glance the stochastic ordering between the 2 samples.

As a standard test for stochastic dominance, we applied the 1-sided 2-sample Kolmogorov-Smirnov test, based on the hypothesis that X has the same distribution as Y against the

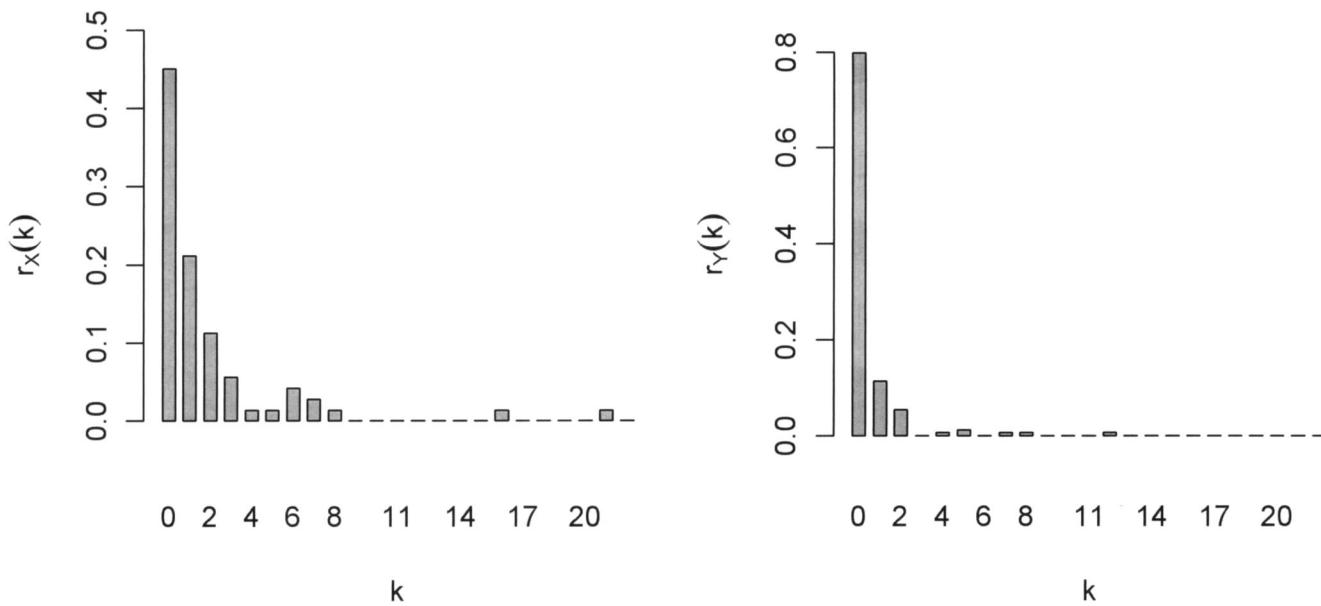


FIGURE 1. Plot of relative frequencies for the 2 samples in example 1.

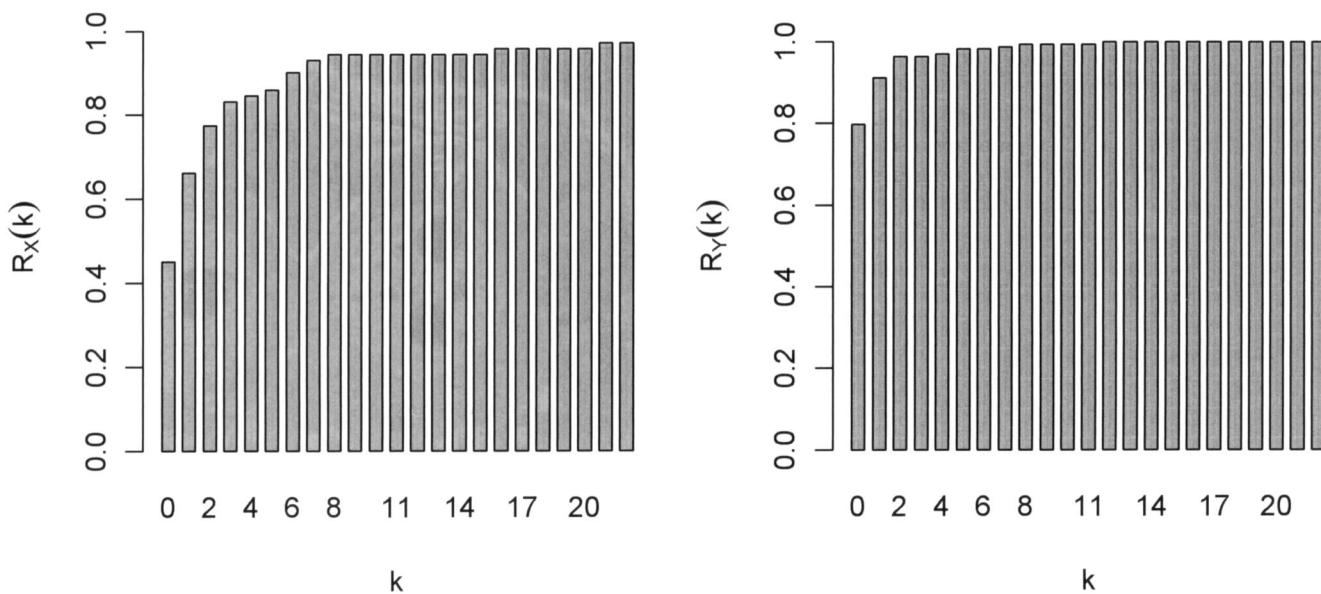


FIGURE 2. Plot of cumulative frequencies for the 2 samples in example 1.

alternative $X \geq_{st} Y$, and leading to a P value less than 0.001. For this test to be exact, the random variables are assumed to be continuous. However, for discrete distributions, the test is still valid, but becomes conservative (Conover, 1999).

TABLE II. Empirical measures for examples 1–3.

Sample sizes	Example 1	Example 2	Example 3
	71 and 168	196 and 100	40 and 20
Diff. between prevalences	0.35	0.08	-0.3
Diff. between means	2.85	-0.05	5.85
Diff. between medians	1	0	-5
Relative effect	0.69	0.53	0.3

FURTHER DATA EXAMPLES

Here we discuss situations in which the 2 samples are not stochastically ordered. A typical situation in which stochastic ordering does not hold occurs when parasite A has a higher prevalence but lower mean (median, etc.) infection intensity than parasite B. That is, if we disregard the uninfected hosts, the distribution of parasite B is stochastically larger (and accordingly, has a higher mean, median, etc.), but if we consider the uninfected hosts as well, then the distributions are no longer stochastically ordered.

In examples 2 and 3, swimbladder and intestinal nematodes in European eels (*Anguilla anguilla*) are considered. Thereby, the differences between the 2 distributions are rather small in the second, and large in the third example.

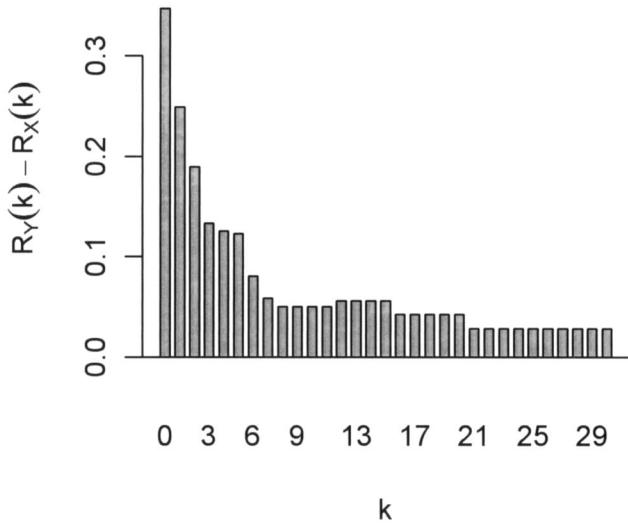


FIGURE 3. P-D plot for the 2 samples in example 1.

As second example, we consider counts of larvae of swim-bladder nematodes in European eels from 2 different locations: sample 1 ($n = 196$) from the River Rhine near Karlsruhe, and sample 2 ($n = 100$) from the River Rhine near Sulzbach (Münsterle, 2005). Again, all recorded nematodes were *A. crassus*. Measures such as the difference between means can be found in the third column of Table II.

The P-D plot (Fig. 5) and the P-P plot (Fig. 6) show that the 2 samples are not stochastically ordered. Because the bars in Figure 5 are rather short, i.e., the differences $R_Y(k) - R_X(k)$ are small, we may assume that both samples come from the same distribution. We conducted a test of the hypotheses $F_X = F_Y$ against the alternative $F_X \neq F_Y$. A suitable test for this purpose is a (2-tailed) 2-sample Kolmogorov-Smirnov test yielding a nonsignificant P value of 0.8. Clearly, this result does not prove the equality of the 2 distributions, but indicates only that the data are compatible with the null hypothesis. One could also apply equivalence tests, which are able to confirm that there are more or

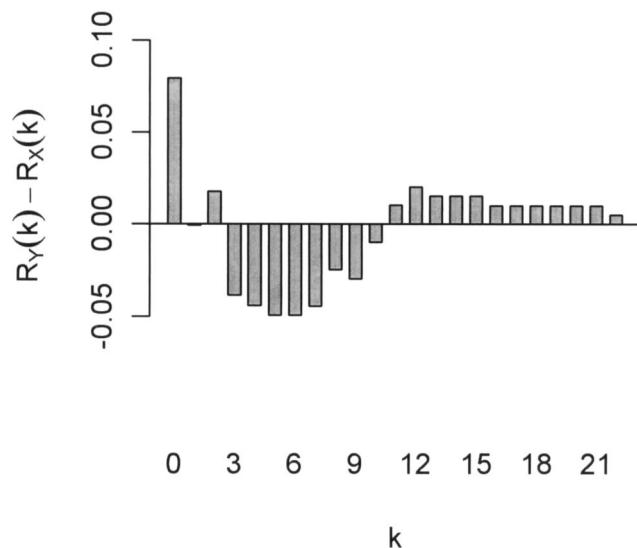


FIGURE 5. P-D plot for the 2 samples in example 2.

less no differences between a selected statistical measure for the 2 samples. Parkhurst (2001), for example, uses the ratio of means, whereas Meier (2009) proposes equivalence tests with respect to the median difference.

The third example compares 2 samples of counts of intestinal parasites in European eels at the same location (River Rhine near Karlsruhe), but from different years. Sample 1 ($n = 40$) was recorded in summer 1999, and sample 2 ($n = 20$) in summer 2005 (Thielen, 2006).

As in example 2, the P-D (Fig. 7) and P-P plots (Fig. 8) show that the 2 samples are not stochastically ordered. However, on examination of the plots, it is quite obvious that the 2 samples do not stem from the same distributions. The prevalence in the second sample is 30% higher than in the first, and the cumulative relative frequencies $R(k)$ are much higher in the second sample for values of k up to 10. The pattern changes for large values of k . Thus, there are noticeably more heavily infected eels in sample 1

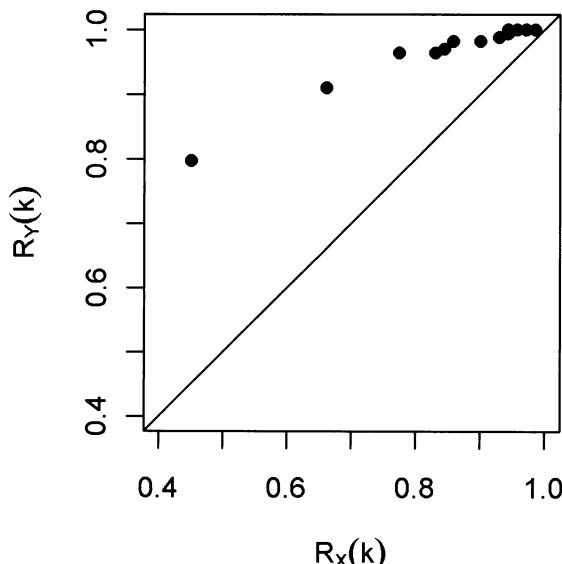


FIGURE 4. P-P plot for the 2 samples in example 1.

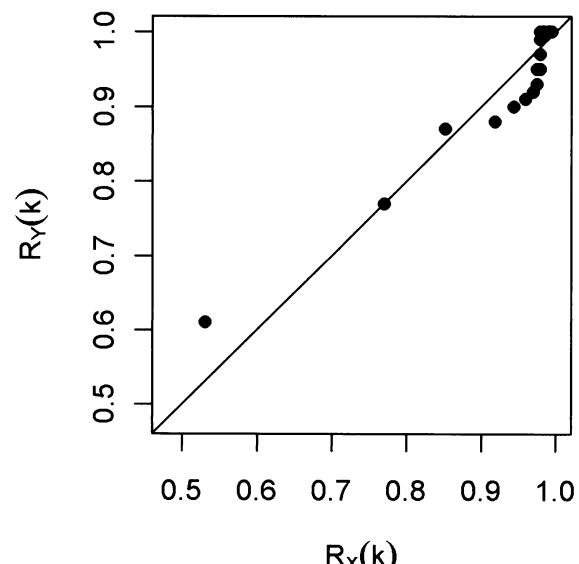


FIGURE 6. P-P plot for the 2 samples in example 2.

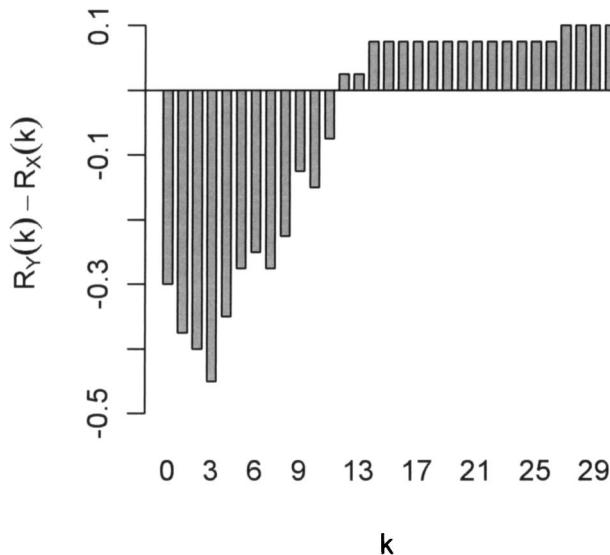


FIGURE 7. P-D plot for the 2 samples in example 3.

than in sample 2. These findings are reflected in the numerical values in the fourth column of Table II.

The reason behind this change in the distribution of intestinal parasites between 1999 and 2005 is explained in Thielen (2006). The parasite *Paratenuisentis ambiguous*, which was predominant in 1999, vanished because of the extinction of an alternate host and was replaced by several other species (*Pomphorhynchus laevis*, *Raphidascaris acus*, *Paraquimperia tenerrima*, *Proteocephalus macrocephalus*, and *Bothrioccephalus claviceps*).

In the present case, it is not generally possible to speak of 1 sample tending to larger values than the other; it is necessary to be more specific: “Sample 1 has a larger mean than sample 2” or “the median is smaller in sample 1,” and we cannot anticipate the ordering when using a 3rd measure.

EXAMPLES OF STOCHASTICALLY ORDERED DISCRETE PROBABILITY DISTRIBUTIONS

The opinion is widespread that stochastic ordering between probability distributions is uncommon. Presumably, this is due to the example of the predominant normal distribution. Two normal distributions can be stochastically ordered only if they have equal variances (and, then, the one with larger mean stochastically dominates the other). However, because equal variance is certainly the exception in practice, the same holds for stochastic ordering between 2 normal distributions. This finding is connected with the fact that the normal distribution can take on both arbitrarily small and large values.

This changes entirely if we consider distributions defined only for positive values such as exponential, gamma, Weibull, or count distributions. We consider the 3 most important count distributions: the Poisson, the binomial, and the negative binomial. Particularly the last distribution is often used to model parasite numbers because data show a high degree of overdispersion (Anderson and Gordon, 1985). As a final example, we consider the logarithmic distribution, which is long and widely used for modeling intensities (see, e.g., Williams, 1964).

First, let X be distributed according to a Poisson distribution with probability density function (pdf) $\lambda^k \exp(-\lambda)/k!$, $k =$

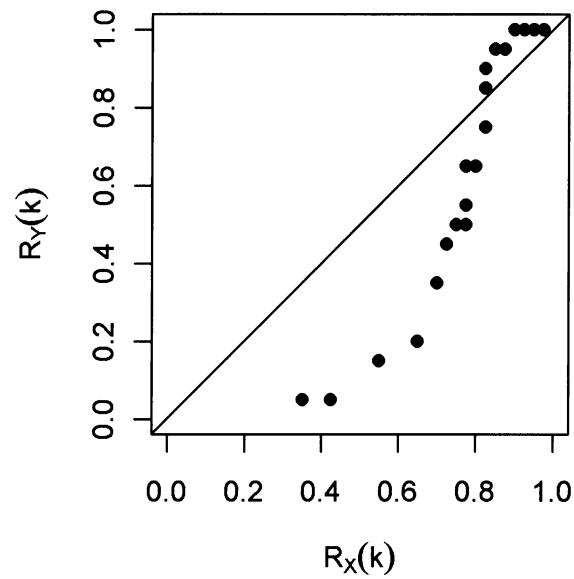


FIGURE 8. P-P plot for the 2 samples in example 3.

0, 1, ..., and let Y be Poisson-distributed with pdf $\mu^k \exp(-\mu)/k!$, $k = 0, 1, 2, \dots$. Then, $X \geq_{st} Y$ if and only if $\lambda \geq \mu$, i.e., if the mean of X is larger or equal to the mean of Y .

Second, let X have a binomial distribution with pdf

$$\binom{m}{k} p^k (1-p)^{m-k}, \quad k = 0, \dots, m,$$

and let Y be binomial distributed with pdf

$$\binom{n}{k} q^k (1-q)^{n-k}, \quad k = 0, \dots, n.$$

Then, $X \geq_{st} Y$ if $p \geq q$ and $m \geq n$, and in this case, $EX \geq EY$ and $Var(X) \geq Var(Y)$.

Third, let X and Y have negative binomial distributions with pdf

$$\binom{k+r-1}{r-1} p^r (1-p)^k, \quad k = 0, 1, 2, \dots,$$

and pdf

$$\binom{k+s-1}{s-1} q^s (1-q)^k, \quad k = 0, 1, 2, \dots,$$

respectively. Then, $X \geq_{st} Y$ if $p \leq q$ and $r \geq s$. In this case, $EX \geq EY$ and $Var(X) \geq Var(Y)$ holds again.

Fourth, let X have a logarithmic distribution with pdf $-\ln(1-p)^{-1} p^k/k$, $k = 1, 2, 3, \dots$, and let Y be logarithmic distributed with pdf $-\ln(1-q)^{-1} q^k/k$, $k = 1, 2, \dots$. Then, $X \geq_{st} Y$ if and only if $p \geq q$.

For the Poisson and the binomial distribution, the results can be found in Müller and Stoyan (2002, table 1.2). Proofs of the claims in case of the negative binomial and logarithmic distribution can be obtained from the authors upon request.

Of course, if we have stochastic ordering in 1 of the 4 examples above, not only are the means and variances ordered, but also

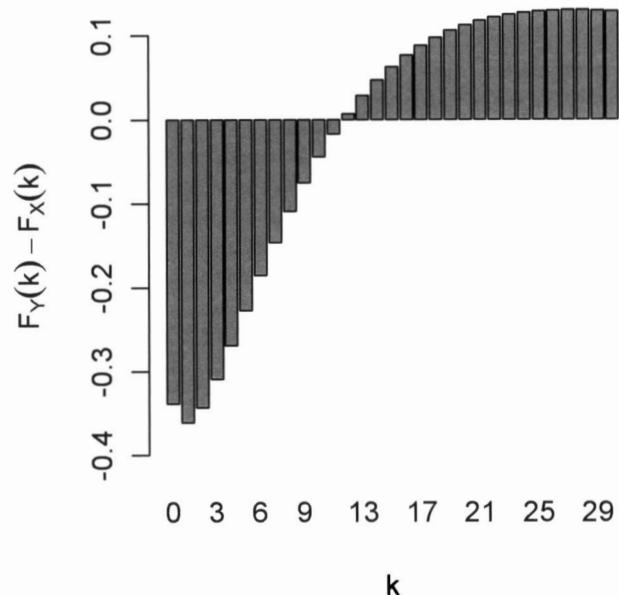


FIGURE 9. P-D plot for the fitted negative binomial distribution for example 3.

medians, etc., as stated above. If, on the other hand, the mean of X is smaller than the mean of Y , but the variances are ordered conversely, both samples will not be stochastically ordered in general.

As examples, we fitted negative binomial distributions to the datasets in examples 1 and 3. This distribution provides a very good model for the data under consideration (see Münsterle et al., 2006 for the first example). In example 1, we obtain $r = 0.28$, $p = 0.08$ for the first sample and $s = 0.17$, $q = 0.27$ for the second sample. Because $r > s$ and $p < q$, we can conclude from the result above that the fitted negative binomial distribution for the first sample stochastically dominates the second, in agreement with the previous findings.

In example 3, we have $r = 0.23$, $p = 0.01$ and $s = 1.60$, $q = 0.15$. Here, $r < s$ and $p < q$, and a P-D plot (Fig. 9) shows that the 2 fitted distributions are not stochastically ordered, which is again consistent with the previous results for this dataset.

CONCLUSION

An important task in many parasitological studies consists of deciding whether the values in one sample tend to be higher than the values of the other sample. However, different dissimilarity measures can lead to different conclusions. A debate as to which measure is superior is unproductive. A similar view with regard to formal tests for the relative effect is expressed in Neuhauser and Ruxton (2009): “Testing the null hypothesis $P = 0.5$ is not necessarily a better or worse approach than comparing central tendencies such as means or medians. However, it offers a

different point-of-view that at least sometimes is more meaningful.”

Using the concept of stochastic dominance, we have proposed graphical tools that will help to identify situations in which it is fully justified to make a general statement that the number of parasites in one sample is larger than in the other sample. Consequently, the most common measures then all lead to the same conclusion. Several examples corroborate the applicability of the concept to empirical data as well as to theoretical count distributions.

ACKNOWLEDGMENTS

The authors thank 2 referees and an associate editor whose comments resulted in a much improved paper.

LITERATURE CITED

ANDERSON, R. M., AND D. M. GORDON. 1982. Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite induced host mortalities. *Parasitology* **85**: 373–398.

BRUNNER, E., AND U. MUNZEL. 2000. The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal* **42**: 17–25.

CONOVER, W. J. 1999. Practical nonparametric statistics. John Wiley and Sons, Chichester, U.K., 592 p.

HOLLANDER, M., AND D. A. WOLFE. 1999. Nonparametric statistical methods. John Wiley and Sons, Chichester, U.K., 787 p.

MEIER, U. 2009. Nonparametric equivalence testing with respect to the median difference. *Pharmaceutical Statistics*. (In press).

MÜLLER, A., AND D. STOYAN. 2002. Comparison methods for stochastic models and risks. John Wiley and Sons, Chichester, U.K., 350 p.

MÜNSTERLE, M. 2005. Ökologische, morphometrische und genetische Untersuchungen an Populationen des invasiven Schwimmblasenematoden *Anguillicola crassus* aus Europa und Taiwan. Ph.D. Dissertation. University of Karlsruhe, Karlsruhe, Germany, 264 p.

—, H. TARASCHEWSKI, B. KLAR, C. W. CHANG, J. C. SHIAO, K. N. SHEN, J. T. HE, S. H. LIN, AND W. N. TZENG. 2006. Occurrence of *Anguillicola crassus* (Nematoda: Dracunculidae) in Japanese eels *Anguilla japonica* from a river and an aquaculture unit in southwest Taiwan. *Diseases of Aquatic Organisms* **71**: 101–108.

NEUBERT, K., AND E. BRUNNER. 2007. A studentized permutation test for the non-parametric Behrens-Fisher problem. *Computational Statistics & Data Analysis* **51**: 5192–5204.

NEUHÄUSER, M., AND R. POULIN. 2004. Comparing parasite numbers between samples of hosts. *Journal of Parasitology* **90**: 689–691.

—, AND G. D. RUXTON. 2009. Distribution-free two-sample comparisons in the case of heterogeneous variances. *Behavioral Ecology and Sociobiology* **63**: 617–623.

PARKHURST, D. F. 2001. Statistical significance tests: Equivalence and reverse tests should reduce misinterpretation. *BioScience* **51**: 1051–1057.

REICZIGEL, J., I. ZAKARIÁS, AND L. RÓZSA. 2005. A bootstrap test of stochastic equality of two populations. *American Statistician* **59**: 156–161.

RÓZSA, L., J. REICZIGEL, AND G. MAJOS. 2000. Quantifying parasites in samples of hosts. *Journal of Parasitology* **86**: 228–232.

THIELEN, F. 2006. Der Einfluss einwandernder Amphipodenarten auf die Parasitozönose des Europäischen Aals (*Anguilla anguilla*). Ph.D. Dissertation. University of Karlsruhe, Karlsruhe, Germany, 256 p.

WILLIAMS, C. B. 1964. Patterns in the balance of nature. Academic Press, London, U.K., 324 p.