The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

Letters to the Editor

Letters to the Editor

FAY, M. P. (2002), "MEASURING A BINARY RESPONSE'S RANGE OF INFLUENCE IN LOGISTIC REGRESSION," *THE AMERICAN STATISTICIAN*, 56, 5–9: COMMENT BY LANDSITTEL, SINGH, ARENA, AND ANDERSON

The recent article by Fay, in which he proposed the range of influence (ROI) statistic for logistic regression, provides a useful diagnostic approach for assessing an observation's "potential influence" on the predicted value (or other statistic of interest). We agree with the author's conclusion that results of this procedure may add substantial information to existing diagnostics. This article was of particular interest to us since we have been investigating the same quantity in a completely different context, namely quantifying degrees of freedom for neural networks and other complex modeling procedures. The primary purpose of this letter is to show how the ROI statistic relates to the concept of "generalized degrees of freedom" as developed by Ye (1998). In the special case of logistic regression, simulation results indicate that the absolute value of the ROI statistic asymptotically corresponds to the diagonal of the hat matrix.

Although the ROI procedure applies to any statistic calculated with logistic regression, $\mathbf{T}(y, x)$, we limit all further discussion to the specific case of $\mathbf{T}_i = \hat{y}_i$. As defined by Fay, the absolute value of the ith ROI statistic is then given by the following expression

$$|\Delta_i(\hat{y}_i)| = |\hat{y}_i - \hat{y}_{i(*i)}|.$$

The term $\hat{y}_{i(*i)}$ represents the ith predicted value obtained by refitting the model to the observed responses, except with y_i switched to $1 - y_i$.

We now discuss the relevance of this statistic to a measure called the "generalized degrees of freedom" (GDF), which Ye proposed for quantifying model complexity in the case of nonparametric models, model selection procedures, and other cases where the degrees of freedom cannot be specified simply as the number of model parameters. Motivated by the definition of the hat matrix for the case of linear regression, where the degrees of freedom

$$p = \text{tr}(\mathbf{H}) = \sum h_{ii} = \sum \partial\hat{y}_i / \partial y_i,$$

Ye conceptualized h_{ii} as the sensitivity of the fitted values to changes in the observed response values. More generally, the previous expression for the degrees of freedom is extended to other modeling procedures by defining the GDF as the $\sum h_i^M(\boldsymbol{\mu})$ where

$$h_i^M(\boldsymbol{\mu}) = \partial E_\mu\left[\hat{\mu}(\mathbf{Y})\right]/\partial\mu_i = \sigma^{-2}\text{cov}\left(\hat{\mu}(\mathbf{Y}), y_i - \mu_i\right),$$

where $Y \sim N(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ and $\hat{\mu}(\mathbf{Y})$ are a set of fitted values from the modeling procedure $\mathbf{M}$. An empirical algorithm is given for estimating this quantity when h_{ii} cannot be explicitly specified. Motivated by Ye's approach, we investigated a modified algorithm for measuring the equivalent of $\partial\hat{y}_i/\partial y_i$, with dichotomous responses.

Unlike linear regression, the hat matrix in logistic regression depends on $\hat{\mathbf{y}}$ and is therefore not constant with respect to changes in $\mathbf{y}$. As referenced by Fay, Pregibon (1981) specified the hat matrix for logistic regression as the following idempotent matrix

$$\mathbf{H} = \mathbf{V}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{1/2},$$

where $\mathbf{V}$ is the diagonal matrix with the ith element equal to $\hat{y}_i(1 - \hat{y}_i)$, and the $\text{tr}(\mathbf{H})$ equals the degrees of freedom. Explicitly solving for h_{ii} is not possible in logistic regression. We proceeded to investigate a dichotomous analog to Ye's GDF with the motivation of approximating $\sum \partial\hat{y}_i/\partial y_i$. For the dichotomous case, each y_i equals 0 or 1, implying that a natural analogy to $\partial\hat{y}_i/\partial y_i$ can be expressed as the increase from $\hat{y}_i$ to $\hat{y}_{i(*i)}$ when y_i equals 0 is replaced by y_i equals 1, and the decrease from $\hat{y}_i$ to $\hat{y}_{i(*i)}$ when y_i equals 1 is replaced by y_i equals 0. Since the ratio $(\hat{y}_i - \hat{y}_{i(*i)})/(y_i - y_{i(*i)})$ is always positive, this expression equates to $|\hat{y}_i - \hat{y}_{i(*i)}|$, which interestingly is the ROI statistic proposed by Fay.

We then investigated the relationship between $|\hat{y}_i - \hat{y}_{i(*i)}|$ and the h_{ii} defined by Pregibon. Since we could not formally quantify the general error introduced, that is,

$$d_i = h_{ii} - |\hat{y}_i - \hat{y}_{i(*i)}|,$$

simulations were conducted to evaluate d_i under various conditions. Both binary and normally distributed covariates were randomly generated, with the response variable randomly generated as a Bernoulli variate, where p_i was calculated by the logistic function of either (i) a constant, (ii) the first covariate only, or (iii) the weighted sum of all covariates. The sample size was varied between 100 to 2,000 (100, 500, and 2,000) and the number of covariates was specified as either 2 or 10. Figure 1 plots the ROI statistic by h_{ii} for a single dataset generated from association (ii), with a sample size of 2,000, and 10 normally distributed covariates. The solid line represents exact concordance between h_{ii} and $|\Delta_i(\hat{y}_i)|$. For this simulation, the $\sum |\Delta_i(\hat{y}_i)|$ was almost exactly equal to the usual degrees of freedom (11.02 versus 11). These results were typical of all other simulatons done with a large sample size relative to the number of covariates (e.g., at least a 50:1 ratio—which includes all simulations done here with two binary or normally distributed covariates in the model). The individual d_i were approximately normally distributed. The total error (i.e., $\sum d_i = \sum h_{ii} - \sum |\Delta_i(\hat{y}_i)|$) for this subset of simulations varied between 0.00 and 0.10. Figure 2 plots the ROI statistic by h_{ii} using a dataset generated from association (ii), a sample size of 100, and 10 normally distributed covariates. For this simulation, which represented the largest discrepancy between $\sum h_{ii}$ and $\sum |\Delta_i(\hat{y}_i)|$, the $\sum |\Delta_i(\hat{y}_i)|$ was 11.29 as compared to 11 degrees of freedom. In the other simulations with 10 covariates and a sample size of 100, the $\sum d_i$ varied between 0.03 and 0.18.

Simulations indicate that the ROI statistic may also yield a useful asymptotic estimate of the h_{ii}. One consequence of our results is that Fay's proposed diagnostic applications for the ROI statistic may be especially informative in cases with relatively small sample sizes, that is, when the ROI statistic and h_{ii} can vary substantially for a given observation. For the simulations with large sample sizes, the ROI statistic (at least approximately) equated to h_{ii}, and thus

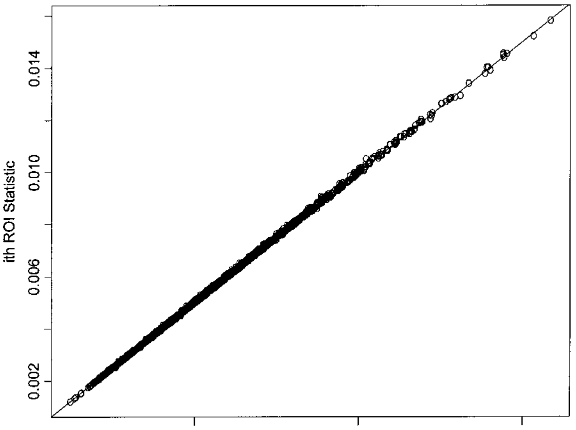

Figure 1. Plot of the ith ROI statistic, $|\Delta_i(\hat{y}_i)|$ by h_{ii} using a simulated dataset with 10 normally distributed covariates, n = 2,000, and a linear association between x_1 and the logit of y. The solid line represents exact concordance between h_{ii} and $|\Delta_i(\hat{y}_i)|$.

 DOI: 10.1198/000313002551

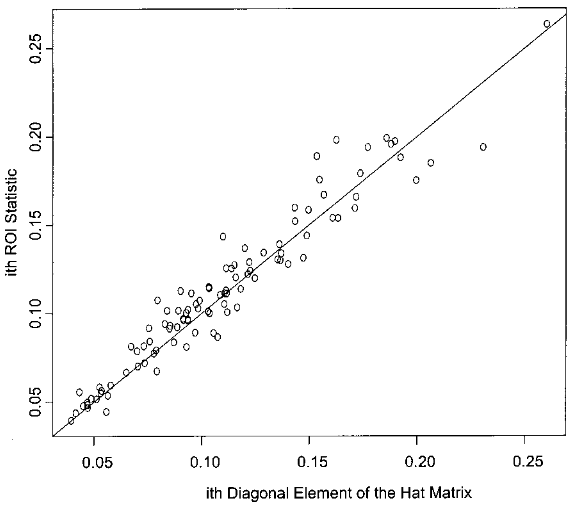

Figure 2. Plot of the ith ROI statistic, $|\Delta_i(\hat{y}_i)|$ by h_{ii} using a simulated dataset with 10 normally distributed covariates, $n = 100$, and a linear association between x_1 and the logit of y. The solid line represents exact concordance between h_{ii} and $|\Delta_i(\hat{y}_i)|$.

provides no additional information. The primary importance of our approach, however, lies in applications outside of generalized linear models where the hat matrix cannot be explicitly specified. Ye discussed at length the application of GDF to such problems (with continuous responses), including nonparametric regression and complex modeling procedures. Our work in extending these results to dichotomous outcomes is motivated by an interest in applying likelihood ratio tests to neural networks. To accomplish this, an appropriate measure must be derived for the degrees of freedom, which does not generally equate to the number of model parameters. Although our initial work on this concept was conducted without knowledge of Fay's study, we wish to thank Dr. Fay for his contributions, as they have enhanced our understanding of the relevant concepts.

REFERENCES

Fay, M. P. (2002), "Measuring a Binary Response's Range of Influence in Logistic Regression," *The American Statistician*, 56, 5–9.

Pregibon, D. (1981), "Logistic Regression Diagnostics," *The Annals of Statistics*, 9, 705–724.

Ye, J. (1998), "On Measuring and Correcting the Effects of Data Mining and Model Selection," *Journal of the American Statistical Association*, 93, 120–131.

Douglas LANDSITTEL
University of Pittsburgh

Harshinder SINGH
West Virginia University
National Institute for Occupational Safety and Health

Vincent C. ARENA
University of Pittsburgh

Stewart J. ANDERSON
University of Pittsburgh

BESSANT, K. C., AND MACPHERSON, E. D. (2002), "THOUGHTS ON THE ORIGINS. CONCEPTS, AND PEDAGOGY OF STATISTICS AS A `SEPARATE DISCIPLINE'," *THE AMERICAN STATISTICIAN*, 56, 22–28: COMMENTS BY LAZAR AND MILLER

LAZAR

In the February 2002 issue of *The American Statistician*, Bessant and MacPherson raised a number of points relating to the history of statistics and perspectives on pedagogy. Their discussion of the shift in mathematics education in the U.S. during the 1950s, to include statistics, brought to mind Moroney's exhortation in his delightful book *Facts from Figures* (Moroney 1967), an introduction to statistics for the general public: "If you are young, then I say: Learn something about statistics as soon as you can. Don't dismiss it through ignorance or because it calls for thought. Don't pass into eternity without having examined these techniques and thought about the possibility of application in your field of work . . . " (p. 463). His final chapter, entitled "Statistics Desirable," also addresses, in a highly entertaining fashion, the question of statistics education in the UK (thus highlighting that the shift was not purely an American phenomenon): "What training do we receive even in the basic ideas of handling statistical data, which is the very lifeblood of everyday life? There is much that can be done with children of even average intelligence before the age of fifteen. But the schoolmasters must learn first . . . . They will be doing a job of inestimable value if they will train their students to have some critical faculty in the face of arithmetic" (p. 459).

Interestingly, Moroney cited the U.S. example to bolster his call for educational reform: "Britain has a proud record in the development of statistics. Yet already a rather familiar story is being retold. Other countries, notably the United States, are more alive to the practical value of these techniques . . . . This book is meant as a small contribution to bringing statistical methods before the attention of those who were raised earlier in history, as well as those who are responsible for the future of our industries" (p. 462).

It is encouraging that, as a discipline, we have made progress since those times, and that the usefulness of statistics to society, and the need for increased numeracy, including early exposure to statistical ideas, have been widely recognized.

Nicole A. LAZAR
Carnegie Mellon University

REFERENCES

Moroney, M. J. (1967), *Facts from Figures* (2nd ed.), Middlesex: Penguin Book.

MILLER

In theory, there is no difference between theory and practice. But, in practice, there is.

—Jan L.A. van de Snepscheut

As a mathematics undergraduate I often looked forward to that portion of the lecture, where the professor might meta-communicate. In other words, he or she would try to give me a glimpse of how the mathematical methods fit into some "big picture" of science. In addition, I always made a mental note when a professor said that a result was important to applications. What took me a while to realize was that this statement could mean different things. Sometimes it meant that a method was of interest to engineers, physicists, or other scientists. On other occasions it meant that the mathematical proof, which we had just seen, would be a useful step in yet another mathematical proof, to be seen later in our studies. This experience made it clearer to me that the way we define our area of study is not a trivial exercise. In fact, those definitions invisibly influence our career paths, and they determine how we will apply what we learn. For this reason I am grateful for the discussion presented by Kenneth Bessant and Eric MacPherson in their article titled "Thoughts on the Origins, Concepts, and Pedagogy of Statistics as a `Separate Discipline'." However, there is one theme to which they allude, namely, the chasm that appears to exist between the theorists and practitioners, which I think deserves to be expanded upon.

A good portion of modern statistics has been concerned with the formal application of probability, and this area of research is closely tied to pure mathematics. A first course in mathematical statistics will make very little reference to data. Much of it will consist of a survey of various probability distributions: gamma, beta, chi-square, Poisson, normal, and others. This is directly comparable to a survey of the various geometries found in mathematics: affine, Euclidean, hyperbolic, projective, and others. The geometer does not present any of these as *the* correct geometry of a given physical phenomenon. He refers that problem to the physicist or some other scientist. However, the geometer can offer these various geometries as a collection of templates, which have been useful in describing physical and abstract spaces. The mathematical statistician can likewise offer the various probability distributions as a collection of templates, which have been useful in describing data. A perfectly straight line cannot be found in the real world, and a perfectly normal distribution cannot be found among real datasets, and yet the concepts of a perfectly straight line and a perfectly normal distribution are useful fictions.

These thoughts suggest how natural it is to view theoretical statistics as a branch of mathematics, employing its power of generalization to connect many seemingly different events and to raise certain aspects of scientific study to a level of abstraction. This power allows a topologist to see the equivalence of the donut and the coffee cup, and it also allows a calculus student to derive a collection of formulas for integrals by using a few mathematical ideas. The power of abstraction, which is so characteristic of mathematics, has been of particular benefit to physicists, who have then been able to condense many different facts into deep and subsuming theories. This situation is very different from that of, say, engineers who apply the principles of physics and other pure sciences. There are, I am sure, general principles that operate in engineering, but there is also so much more in the way of combining and weaving of principles from various sciences, plus the gleaning of miscellaneous facts and special cases found by empirical methods, all of which bring into focus the issue of *context*—that is, how general principles are to be applied in varying real situations. This is the world of the engineer, and it is certainly the world of the applied statistician. It is a much different world from that of mathematics, which can sometimes seem to the outsider to be, as Hermann Weyl said, "a logician's paradise . . . endowed with the ultimate furniture" (Kline 1980). Although they are indispensable tools of science, mathematics and theoretical statistics might not be considered sciences in themselves, because they do not rely upon any tests of experience.

Holmes (1985) quoted the report of the Cockcroft Committee, which studied mathematical education in Britain: ". . . statistics is not just a set of techniques, it is an attitude of mind approaching data." Just where does this "attitude of mind" come from? Since data are only of passing interest to the theorist, it clearly does not come from mathematical statistics. Although there is not yet any standard notion of what the phrase "statistical thinking" means, intuitively it seems to refer to some things which are not usually taught at universities, but instead are learned through experience. The fact that statisticians have created a phrase to describe this knowledge might indicate that our field has a greater-than-average divide between what is taught in school and what is used through experience. From my own experience there certainly does exist an attitude that develops gradually from a long steady diet of data analysis, which influences our sense about how general principles are to be applied, and which alters our belief about the nature of inductive reasoning and human bias. The sheer variety of datasets is much greater than can be anticipated by a study of theory, and this probably prevents us, while we are at universities, from seeing how fallible we can be when confronted with complicated and large datasets. These experiences take their place in our consciousness, and become the true priors of an analysis. For myself I believe that statistical thinking could include the following:

1. Modeling and formal methods are not always the key to an analysis. For some studies, the most difficult and crucial step may consist of understanding just how the data was collected. The description of data seems to be less important in theory than in practice, where the hallmark of a persuasive analysis might consist of how the descriptive and formal components of the analysis fit together. One aspect of this is that, as Tufte (1997) suggested, "the logic of display design must reflect the logic of the analysis." This can be difficult to achieve in practice.

2. Practitioners need to be generalists, familiar with the properties of many different methods without necessarily understanding the precise theory of all of them. Theorists or specialists can be so fond of their particular methods that they will sometimes fit the data to the methods and not the other way around.

3. As Mallows (1998) said, "The main challenge of applied statistical work is that of taking proper account of contextual issues." One implication of this is that applied work is about making decisions: How should we amend a standard method to account for the limitations of the data? What criteria should we use for excluding data? How do we define qualitative differences? How do we determine how many hypotheses a dataset can support? How do we decide what compromises to allow between theory and practice? A sign of naive research is to take refuge in a complicated model or to expect a complicated procedure to always make the hard decisions for you. In addition, increasing the mathematical complexities of an analysis will not always lead to significant gains in understanding.

4. As Tukey and others have suggested, the more p values you generate during an analysis, the less they mean. But, even when few p values are generated, their meaning is not necessarily clear. As Hahn and Meeker (1993) have written, once we have considered our assumptions, and also have accounted for the differences between the target population and the sampling frame, and between some latent variable and the actual measurement, what we are left with in a p value or confidence interval may only be an indication of the quality of the information, not some absolute criteria for decision-making.

5. Distinctions that are important to theory may not always have the same relevance in practice. For instance, it is possible that a misunderstanding of the theoretical difference between frequentist and Bayesian approaches to probability will undermine an analysis, but it is doubtful that in practice it would have the same consequences as misunderstanding the difference between the results of controlled experiments, where precise questions can be posed and answered, and the results of observational studies, where selection bias and confounding can often dominate the conclusions. [I did not fully appreciate this distinction until long after I left the university. In retrospect I wish that the article by Box (1966) on the use and abuse of regression methods had been required reading during my graduate studies.]

6. Technical understanding is not enough; some conceptual understanding is necessary. Even an old idea can have modern applications. For instance, suppose that a researcher is interested in detecting a mean change of at least 10% from the first to the second responses, but only for the fifth percentile of the first responses from a group of subjects. In that case any use of within-subject contrasts will have to account for the effects of regression towards the mean.

7. Many of the issues, with which applied statisticians must deal, are not strictly statistical in nature, and there are many different roles that they must play as they apply their craft. For instance, a statistician sometimes has the important, and often thankless, role of providing some caution, not only on the direction, but also the scale, of the results. This is especially important in an age of competitive funding, when negative studies are generally more difficult to get published, attract less attention and are less read, and when awards are rarely bestowed on researchers who, for instance, find that there is no association between a certain exposure and disease. In the light of these realities, it is no wonder that researchers are sometimes distressed by negative results, or wish to ignore the distinction between confirmatory and exploratory methods of analysis. Bias can at times be evident in the interpretation of the results from some studies, such as when causality is confused with correlation. However, it is the applied statistician who is best able to resist the effect that bias can have on the invisible day-to-day process of data analysis.

During my initial years of consultation it was interesting to me to find that the researchers who had had the most previous coursework in statistics, would also be the ones who at times expressed a lack of confidence in their ability to apply statistics. It seems to me that they were discovering that the process of data analysis involves much more than choosing the most appropriate statistical test. Best (1992) and others have tried to describe this process by identifying some components of it, such as (a) understanding the goals of the analysis, (b) understanding the data structure(s), (c) describing the data, (d) undertaking the formal analysis, and (e) interpreting and communicating the results. The complexity of the process can be illustrated by considering the falsity of a maxim referred to by Hoerl, Hooper, Jacobs, and Lucas (1993): "If you know how to use a hammer, saw, level, and ruler, you will be able to build a house." In the case of data analysis, knowing the individual properties for an array of tools does not necessarily mean that you will be confident about how the tools work together. We know that certain tools can be complementary, like cluster analysis and multidimensional scaling, but there seems to be very little knowledge about this in general. In addition, statistical methods rarely come with any advice about user variation.

Mallows (1998) stated that, “We have almost no theory to help us understand how to think about applied statistics.” This could imply that, in the coming decades, data analysis will take on some very different appearances, which will ultimately make it as different a subject from theoretical statistics as engineering is different from physics. Theoretical statistics might not be considered a science in the strict sense of the word, but it seems to me that data analysis is destined to be a science in every sense of the word. The practice of data analysis does, indeed, rely on tests of experience with data. Some initial evidence for this can be found in the use of experimental methods to show, for instance, that the perception of correlation is dependent on the scale of a scatterplot (Cleveland 1982), or to show that certain attributes, such as differences in length and position, are easier to perceive in plots than other attributes, such as differences in area or angle (Cleveland and McGill 1985). The systematic use of statistical methods is being turned upon itself, and this process could eventually provide some deeper insights into aspects of data analysis. For example, Hoaglin and Velleman (1995) compared 15 different analyses of a large baseball dataset. They noticed that it was more difficult for some of the data analysts to find the multivariate outliers if they did not first identify and remove the univariate outliers. When should we identify and remove univariate outliers before we search for multivariate outliers? There are no clear answers for this critical question, and yet there is no reason to believe that empirical approaches couldn't eventually provide some insight by using representative samples of methods and datasets. Perhaps, it will also be possible to bench-test the user variation for a new method in the same way that we now bench-test a new statistical package for accuracy.

A consulting statistician might be viewed as somewhat of a dilettante in our age of specialists. It is also a more difficult time to be a consulting generalist than a university specialist. For instance, consider one of the more formidable problems in data analysis: how to uncover hidden structures in data, such as homogeneous subgroups, correlations that only exist for a special subset of observations, or blocks of erroneous data entries. Graphs are good for low dimensions, but what should one use in higher dimensions? New methods are proliferating, but without any clear sense of how much more efficient one is over the other. If you are a specialist in one of many areas—projection pursuit, data mining techniques, grand tours, multidimensional scaling, cluster analysis, factor analysis, correspondence analysis, outlier analysis, and so on—then your energies can be channeled into promoting the advantages of your tool of choice. However, if you are a generalist trying to find the best possible method or combination of methods, you may ultimately experience a paralyzing moment when too many choices and no choices become one.

I believe that statistics is comprised of two very different but complementary branches, whose respective members do not always understand each others' perspectives. For example, communication and developing trust are extremely important in consulting work. Theoretical statisticians may not always appreciate this, and have in some cases inoculated researchers against any further statistical advice. In fact, some theoreticians should probably stay away from consulting work, but I don't mean to imply any shame in this statement. Consider the medical student who does not particularly enjoy clinical work, but instead enters into research. Doesn't medicine need both applied and theoretical scientists? Tukey (1997) stated that, “data analysis, to be practical, has to be engineering, not science.” To achieve this, I believe that statistics will need both applied and theoretical statisticians, who work together the way that engineers and physicists might ideally work together. It is not helpful to promote the fiction that the theory and practice of statistics are the same. As someone who is working in the trenches, I feel that we are sometimes fighting a losing battle against the prevailing notion in some circles of science that statistics is no more than a machine that converts data into truth. Promoting the idea that the application of statistics also has many subtleties, which are distinct from the subtleties of theory, will help considerably to dispel that notion.

William E. MILLER
Morgantown, WV

REFERENCES

Bessant, K.C., and MacPherson, E. D. (2002), “Thoughts on the Origins, Concepts, and Pedagogy of Statistics as a `Separate Discipline',” *The American Statistician*, 56, 22–28.

Best, A. (1992), “Interactive Data Analysis Strategies,” in *SAS Users Group International Conference Proceedings*, Cary, NC: SAS Institute, pp. 1417–1422.

Box, G. E. P. (1966), “Use and Abuse of Regression,” *Technometrics*, 8, 625–629.

Cleveland, W. S. (1982), “Variables in Scatterplots Look More Correlated When the Scales Are Increased,” *Science*, 216, 1138–1141.

Cleveland, W. S., and McGill, R. (1985), “Graphical Perception and Graphical Methods for Analyzing and Presenting Scientific Data,” *Science*, 229, 828–833.

Cockcroft, W. H. (1982), *Mathematics Counts. Report by the Inquiry into the Teaching of Mathematics in Schools*, London: HMSO.

Hahn, G. J., and Meeker, W. Q. (1993), “Assumptions for Statistical Inference,” *The American Statistician*, 47, 1–11.

Hoaglin, D. C., and Velleman, P. F. (1995), “A Critical Look at Some Analyses of Major League Baseball Salaries,” *The American Statistician*, 49, 277–285.

Hoerl, R. W., Hooper, J. H., Jacobs, P. J., and Lucas, J. M. (1993), “Skills for Industrial Statisticians to Survive and Prosper in the Emerging Quality Environment,” *The American Statistician*, 47, 280–292.

Holmes, P. (1985), Commentary on “The Initial Examination of Data” by C. Chatfield, *Journal of the Royal Statistical Society*, Ser. A, 148, 233–234.

Kline, M. (1980), *Mathematics: The Loss of Certainty*, New York: Oxford University Press.

Mallows, C. (1998), “The Zeroth Problem,” *The American Statistician*, 52, 1–9.

Tufte, E.R. (1997), *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, CT: Graphics Press.

Tukey, J. (1997), “More Honest Foundations for Data Analysis,” *Journal of Statistical Planning and Inference*, 57, 21–28.

VAN DORP, J. R., AND KOTZ, S. (2002), “THE STANDARD TWO-SIDED POWER DISTRIBUTION AND ITS PROPERTIES: WITH APPLICATIONS IN FINANCIAL ENGINEERING,” *THE AMERICAN STATISTICIAN*, 56, 90–99: COMMENT BY NADARAJAH AND REPLY

Van Dorp and Kotz (2002) introduced a two-sided power distribution with the probability density function

$$f(x|\theta, n) = \begin{cases} n(\frac{x}{\theta})^{n-1}, & \text{if} \quad 0 < x \le \theta, \\ n\left(\frac{1-x}{1-\theta}\right)^{n-1}, & \text{if} \quad \theta \le x < 1, \end{cases} \tag{1}$$

parameterized by $n > 0$ and $0 \le \theta \le 1$. This distribution contains as special cases the triangular distribution, the standard power function distribution, and the uniform distribution. The authors derived various properties of (1) and discussed its flexibility as compared with the beta family. They also provided a novel application in the area of financial engineering.

In this note I would like to point out that (1) can be reformulated in such a way—without adding more parameters—that the resulting distribution is more flexible. For $0 < \theta < 1$, define

$$f(x|\theta, n) = \begin{cases} cx^{m-1}, & \text{if} \quad 0 < x \le \theta, \\ c(1-x)^{n-1}, & \text{if} \quad \theta \le x < 1. \end{cases} \tag{2}$$

Here, $m > 0$ and $n > 0$ are chosen so that f is continuous at θ, that is, $\theta^{m-1} = (1-\theta)^{n-1}$ and c is the proportionality factor

$$c = 1 \Big/ \left\{ \frac{\theta^m}{m} + \frac{(1-\theta)^n}{n} \right\}$$

to make sure that (2) integrates to 1. This formulation is similar to the one adopted by Nadarajah (1999). In this formulation there is no need for the powers m and n to be equal. In fact, $m = n$ if and only if $\theta = 1/2$ or $m = n = 1$. But if $m > 1$, then also $n > 1$ and if $m < 1$, then also $n < 1$. Like (1), (2) has only two free parameters and contains the triangular distribution, the standard power function distribution, and the uniform distribution as special cases. Standard calculations show that the cumulative distribution function (cdf) associated with (2) is:

$$F(x|\theta, n) = \begin{cases} \frac{cx^m}{m}, & \text{if} \quad 0 < x \le \theta, \\ 1 - \frac{c(1-x)^n}{n}, & \text{if} \quad \theta \le x < 1 \end{cases} \tag{3}$$

and the kth moment is:

$$E\left(X^k\right) = \frac{c\theta^{k+m}}{k+m} + c\sum_{i=0}^{k}(-1)^i \binom{k}{i}\frac{(1-\theta)^{n+i}}{n+i}.$$

To illustrate the flexibility of (2) over (1), consider the table below where I have computed numerically the range of possible values of $E(X)$ and var(X) for the two models for selected values of n.

	$E(X)$		var(X)	
n	Model (1)	Model (2)	Model (1)	Model (2)
0.1	(0.091, 0.909)	(0.000, 0.909)	(0.039, 0.216)	(0.000, 0216)
1	(0.5, 0.5)	(0.5, 0.5)	(0.083, 0.083)	(0.083, 0.083)
5	(0.167, 0.833)	(0.167, 0.999)	(0.012, 0.020)	(0.000, 0.020)
10	(0.091, 0.909)	(0.091, 0.999)	(0.004, 0.007)	(0.000, 0.007)

Evidently the reformulated model (2) provides a wider range of values for $E(X)$ and var(X) for each n. Another evidence of flexibility of (2) over (1) can be obtained by drawing the moment ratio diagram for (2) and comparing it with the corresponding figure for model (1), Figure 5 in van Dorp and Kotz (2002).

Like in Equations (8) and (9) of van Dorp and Kotz (2002), the cdf (3) for (2) has the stochastic increasing property $\theta_1 < \theta_2, x \in (0,1) \Longrightarrow F(x|\theta_1, n) > F(x|\theta_2, n)$ when $n > 1$ and has the stochastic decreasing property $\theta_1 < \theta_2, x \in (0,1) \Longrightarrow F(x|\theta_1, n) < F(x|\theta_2, n)$ when $n < 1$. The proof of this statement, which I shall not discuss, is algebraically more demanding. The pth percentile x_p defined by $F(x_p \mid \theta, n) = p$ satisfies properties similar to those discussed in Section 2.3 of van Dorp and Kotz (2002). For instance, $x_p < \theta$ if and only if $p < (n\theta)/\{n\theta + m(1-\theta)\}$ and $x_p = \theta$ if and only if $p = (n\theta)/\{n\theta+m(1-\theta)\}$. The limiting distribution of $Y = (n-1)A(X-\theta)/\theta$ as $n \to \infty$ approaches an asymmetric Laplace distribution like in Equation (19) of van Dorp and Kotz (2002). But the expression for the relative entropy turns out to be:

$$E[-\log f(X)] = -\log c + \frac{c(n-1)}{n^2}(1-\theta)^n - \frac{c(n-1)}{n}(1-\theta)^n \log\theta + \frac{c(m-1)}{m^2}\theta^m - \frac{c(m-1)}{m}\theta^m \log\theta,$$

which is a function of both n and θ. The relative entropy for (1) in Equation (23) of van Dorp and Kotz (2002) depends only on n.

Saralees NADARAJAH
University of South Florida

REFERENCES

Nadarajah, S. (1999), "A Polynomial Model for Bivariate Extreme Value Distributions," *Statistics and Probability Letters*, 42, 15–25.

van Dorp, J. R., and Kotz, S. (2002), "The Standard Two-Sided Power Distribution and its Properties: With Applications in Financial Engineering," *The American Statistician*, 56, 90–99.

REPLY

We were delighted to learn details about the clever reparameterization, provided by Professor S. Nadarajah of the standard two-sided power (STSP) distribution presented by van Dorp and Kotz (2002). It is indeed encouraging that the ranges of $E[X]$, var(X), and the moment ratio diagram involving $\sqrt{\beta_1}$ and β_2 are wider in his reparameterization, which further enhances the flexible STSP distribution proposed by us. However, we fail to see the transparency (or the meaning) of Nadarajah's parameters; in our opinion this is a crucial aspect for practical applications.

Perhaps this may be illustrated by the fact that in the original representation (3) and (4) of van Dorp and Kotz (2002) the parameter θ is simply the θth percentile, that is, $F(\theta|\theta, n) = \theta$, and the relative entropy is $\log n - \frac{n}{n-1}$. In Nadarajah's reparameterization θ corresponds to the $c\theta^{m-1}$th percentile, where

$$m = 1 + (n-1)\frac{\log(1-\theta)}{\log\theta},$$

$$c = 1 \Big/ \left\{\frac{\theta^m}{m} + \frac{(1-\theta)^n}{n}\right\},$$

and the relative entropy is given by a rather complicated expression also involving all four parameters θ, n, c, and m. We are also somewhat concerned about the structure of the expression for $E[X|\theta, n]$, the moment and (especially) the MLE estimations in Nadarajah's reparameterization of the STSP distribution and would be happy to learn more about these points.

In summary, the STSP distribution in Van Dorp and Kotz (2002) was devised as an alternative to the widely used beta distribution—specifically when the properties of the STSP distribution complement those of the beta distribution. Examples of these properties are: the meaning of its parameters, the intuitive structure of the expression for $E[X|\theta, n]$ as a weighted average of its support boundaries 0 and 1 and of θ, that is, $E[X|\theta, n] = \frac{0+(n-1)\theta+1}{n+1}$, an MLE procedure which involves only elementary functions, and a transparent form for its entropy function. It would seem that some of these advantages are being lost in the reparameterization of the STSP distributions suggested by Professor S. Nadarajah.

J. R. VAN DORP
S. KOTZ
The George Washington University

BARKER, L. (2002), "A COMPARISON OF NINE CONFIDENCE INTERVALS FOR A POISSON PARAMETER WHEN THE EXPECTED NUMBER OF EVENTS IS ≤ 5," *THE AMERICAN STATISTICIAN*, 56, 85–89: COMMENT BY THÖNI

I would like to draw your attention to the final remark referring to method 9: The Exact Method (E).

A formal solution of the equations given for method E can be found using a relation between the chi-squared and the Poisson distributions (e.g., Pearson and Hartley 1970, p. 128), and the lower and upper limits of the confidence interval for the Poisson parameter can be computed in closed form as a function of lower and upper quantiles of the chi-squared distribution, the number of observed events, and the sample size (Johnson, Kotz, and Kemp 1992, p. 171).

Since modern programmable pocket calculators have algorithms to compute the quantiles of the chi-squared distribution, the exact confidence limits for Poisson parameters according to method E can be very easily programmed as a function of the observed number of events and the sample size.

Using the same notation as Barker, the lower and upper limits of the exact $1-\alpha$ confidence interval for λ can be written to be

$$\lambda_l = \frac{1}{2n}\chi^2_{2n\overline{X};1-\alpha/2}; \quad \lambda_u = \frac{1}{2n}\chi^2_{2n\overline{X}+2;\alpha/2},$$

where $\chi^2_{v;Q}$ is the $(1-Q)$ 100th percentile of the chi-squared distribution with v degrees of freedom.

Hanspeter THÖNI
University at Hohenheim
Stuttgart, Germany

REFERENCES

Pearson, E. S., and Hartley, H. O. (1970), *Biometrika Tables for Statisticians* (vol. I), New York: Cambridge University Press.

Johnson, N. L., Kotz, S., and Kemp, A. W. (1992), *Univariate Discrete Distributions* (2nd ed.), New York: Wiley.

LANGFORD, E., SCHWERTMAN, N., AND OWENS, M. (2001), "IS THE PROPERTY OF BEING POSITIVELY CORRELATED TRANSITIVE?", *THE AMERICAN STATISTICIAN*, 55, 322–325: COMMENT BY LIPOVETSKY

The authors showed that for three random variables X, Y, and Z, if Y is positively correlated with X and with Z, the latter two can be correlated negatively. It is a nice article and I would like to make just a couple of comments.

In Section 1, the authors gave an example (and indicate that the reasoning is false) that for positive regression trends of Y by X and of Z by Y it is easy to conclude that the trend of Z by X should be positive, too. Indeed, suppose for the standardized variables we constructed the models $Y = aX$ and $Z = bY$, where coefficients of regressions correspond to correlations $a = \rho_{XY}$ and $b = \rho_{YZ}$. Substituting the former model into the latter yields a model $Z = cX$ with the coefficient $c = ab = \rho_{XY}\rho_{YZ}$. By this product it seems evident that for both positive trends the combined trend should also be positive. However, it is not necessarily so, and the easy way to see it is as following. If we want to use Y expressed via X in the model for Z, we should take not the empirical values of Y but its theoretical values $\hat{Y}$ estimated by the model $\hat{Y} = aX$. Then in the model $Z = c\hat{Y}$ the coefficient of regression equals

$$c = \frac{\operatorname{cov}(Z, \hat{Y})}{\operatorname{var}(\hat{Y})} = \frac{a \operatorname{cov}(Z, X)}{a^2 \operatorname{var}(X)} = \frac{\operatorname{cov}(Z, X)}{a} = \frac{\rho_{XZ}}{\rho_{XY}}. \tag{1}$$

So we can see that the trend is defined not just by positive ρ_{XY} but also by the sign of the correlation ρ_{XZ} of Z directly with X.

In Theorem 2 (p. 323) it is shown that for correlations $\rho_{XY} > 0$ and $\rho_{YZ} > 0$, if

$$\rho_{XY}^2 + \rho_{YZ}^2 > 1, \tag{2}$$

then the third correlation is positive, $\rho_{XZ} > 0$. It is correct but can be easily generalized saying that for both positive $\rho_{XY} > 0$ and $\rho_{YZ} > 0$, or both negative correlations $\rho_{XY} < 0$ and $\rho_{YZ} < 0$, the same condition (2) is sufficient for the third correlation to be positive, $\rho_{XZ} > 0$. We can also say that if one correlation is positive, $\rho_{XY} > 0$, and another is negative, $\rho_{YZ} < 0$, then condition (2) is sufficient to have the third correlation negative, $\rho_{XZ} < 0$. So this inequality (2) defines transitive sets of three correlations with signs $(+++)$ or $(--+)$ corresponded to the "deterministic" signs of two numbers and their product in regular arithmetic.

Stan LIPOVETSKY
Custom Research, Inc.
Minneapolis, MN