Statistical Science
2005, Vol. 20, No. 4, 382–383
DOI 10.1214/088342305000000386

Comment: Fuzzy and Randomized Confidence Intervals and P-Values

Elizabeth A. Thompson

Geyer and Meeden are to be congratulated for a major idea on how to express uncertainty in classical frequentist statistics. Their development of the trinity of fuzzy test functions, confidence intervals and P-values brings a new coherence to the relationship among these statistical entities when the test is randomized. In Geyer and Meeden, the uncertainty or randomization is due to the discreteness of data random variables, but another form of uncertainty or randomness arises in latent variable problems and may be treated in an analogous fashion. In this discussion, I will focus on fuzzy P-values in that context.

Following the notation of Geyer and Meeden, let $\mathbf{X}$ denote the observed data random variables, but suppose there are latent variables $\mathbf{W}$ of scientific interest, so that the ideal test statistic is $t(\mathbf{X}, \mathbf{W})$, a function of both $\mathbf{X}$ and $\mathbf{W}$. Many examples of this situation arise in the analysis of genetic data, where $\mathbf{W}$ may be unobservable DNA types or paths of descent of DNA to the observed individuals of a pedigree or population. Indeed, in this case the hypothesis of interest often concerns the probability distribution of $\mathbf{W}$ and hence the ideal test statistic is a function of $\mathbf{W}$ alone. The data random variable $\mathbf{X}$ is only of interest for the information it provides about $\mathbf{W}$ through some probability model $\Pr(\mathbf{X}|\mathbf{W})$. For convenience, we consider here the case where the latent test statistic $t(\mathbf{W})$ is a function only of $\mathbf{W}$ and assume the random variable $\mathbf{W}$ to be continuous. The discrete case is considered by Thompson and Geyer (2005).

In the area of statistical genetic methodology, a standard procedure has been to average over the latent-variable uncertainty and form test statistics $\mathrm{E}(t(\mathbf{W})|\mathbf{X})$ (Whittemore and Halpern, 1994; Kruglyak, Daly, Reeve-Daly and Lander, 1996). However, there appear to be neither theoretical justification nor optimality properties for such a proceeding. Moreover, the distribution of such a test statistic is not only hard or impossible to obtain, but depends on the distribution of $\mathbf{X}$ given $\mathbf{W}$, which may itself be subject to considerable uncertainty. The second key point made by Geyer and Meeden is that the distribution function of the abstract randomized (fuzzy) P-value that results from an analysis of data should be reported *without additional and arbitrary randomization*. Likewise, in the latent variable case, a solution other than averaging over the uncertainty in $t(\mathbf{W})$ given $\mathbf{X}$ is clearly desirable.

Thompson and Basu (2003) first attempted to address this problem by considering the distribution over $\mathbf{X}_0$ of probabilities of the form

$$(1) \qquad Q(\mathbf{X}, \mathbf{X}_0) = \Pr\big(t(\mathbf{W}) > t(\mathbf{W}_0)|\mathbf{X}, \mathbf{X}_0\big),$$

where $\mathbf{X}_0$ is independent of $\mathbf{X}$ and has the distribution of $\mathbf{X}$ under the null hypothesis, $\mathbf{W}_0$ is independent of $\mathbf{W}$, and the distribution of $\mathbf{W}_0$ given $\mathbf{X}_0$ is that of $\mathbf{W}$ given $\mathbf{X}$ under the null hypothesis. Under the null hypothesis, the distribution of $Q(\mathbf{X}, \mathbf{X}_0)$ is symmetric about $1/2$. Furthermore, if the structure of the data is such that $\mathbf{X}$ ($\mathbf{X}_0$) leaves little uncertainty in $t(\mathbf{W})$ [$t(\mathbf{W}_0)$], the distribution of the function $Q(\mathbf{X}, \mathbf{X}_0)$ of the random variable $\mathbf{X}_0$ will take extreme values. In the limit, where $t(\mathbf{W})$ and $t(\mathbf{W}_0)$ are deterministic functions of $\mathbf{X}$ and $\mathbf{X}_0$, respectively, the distribution of Q has two-point support $\{0, 1\}$. At the other extreme, where $\mathbf{X}$ and $\mathbf{X}_0$ provide no information about $\mathbf{W}$ and $\mathbf{W}_0$, Q is identically equal to $1/2$.

The probability distribution of $Q(\mathbf{X}, \mathbf{X}_0)$ over $\mathbf{X}_0$ for the given data $\mathbf{X}$ goes some way toward differentiating the evidence $\mathbf{X}$ provides about $\mathbf{W}$ from the evidence $\mathbf{W}$ provides about the null hypothesis. However, the construction of (1) has several disadvantages. The distributions must normally be estimated by Monte Carlo methods, and realizations of $\mathbf{X}_0$ and then of $\mathbf{W}_0$ for each realized $\mathbf{X}_0$ are required. The result is clearly dependent on the assumed model for $\mathbf{X}_0$. Furthermore, rigorous interpretation of the distributions of Q is unclear. As seen from the two extreme cases of the previous paragraph, Q is not a P-value: it does not have a uniform distribution on $[0, 1]$ when $\mathbf{X}$ is generated under the null hypothesis.

The approach of Geyer and Meeden provides a more satisfactory solution. With the same notation as above,

Elizabeth A. Thompson is Professor, Department of Statistics, University of Washington, Seattle, Washington 98195-4322, USA (e-mail: thompson@stat.washington.edu).

let

$$s(\mathbf{w}) = \Pr\{t(\mathbf{W}_0) \geq t(\mathbf{w})\}$$

be the P-value of the test where the value $\mathbf{w}$ of $\mathbf{W}$ is to be observed. We refer to $s(\mathbf{W})$ as the *latent P-value*. Equivalently, we may write

$$s(\mathbf{W}) = \Pr\{t(\mathbf{W}_0) \geq t(\mathbf{W}) | \mathbf{W}\}. \tag{2}$$

The random variable $s(\mathbf{W})$ is the fuzzy P-value and has distribution function

$$F(\alpha) = \Pr\{s(\mathbf{W}) \leq \alpha | \mathbf{X}\}. \tag{3}$$

Exactly as in (1.6) and (1.7) of Geyer and Meeden, iterated conditional expectation provides that, under H_0, the unconditional expectation of $F(\alpha)$ is α.

Note that the (distribution of the) fuzzy P-value is a function of the observed data $\mathbf{X}$. The distribution of $\mathbf{X}$ under the null hypothesis (or $\mathbf{X}_0$) does not enter into the computation. This not only greatly facilitates computation, but also provides robustness to assumptions with regard to the marginal distribution of $\mathbf{X}_0$.

Finally, we return again to the telling plea of Geyer and Meeden that the fuzzy P-value distribution be provided by data analysts, *without additional randomization*. In the latent variable case, a classical P-value confounds the evidence in data $\mathbf{X}$ about the latent variables $\mathbf{W}$ with the evidence that $\mathbf{X}$ [through $t(\mathbf{W})$] provides about the hypothesis of interest. Like the probabilities of (1), the fuzzy P-value of (2) separates the evidence that $\mathbf{X}$ provides about the latent test statistic $t(\mathbf{W})$ from the evidence that the data provide for inference. Unlike the probabilities of (1), the fuzzy P-values of (2) are rigorously interpretable as P-values that have a uniform distribution under the null hypothesis (3). Clearly, the latter provide a better solution.

REFERENCES

KRUGLYAK, L., DALY, M. J., REEVE-DALY, M. P. and LANDER, E. S. (1996). Parametric and nonparametric linkage analysis: A unified multipoint approach. *American J. Human Genetics* **58** 1347–1363.

THOMPSON, E. A. and BASU, S. (2003). Genome sharing in large pedigrees: Multiple imputation of *ibd* for linkage detection. *Human Heredity* **56** 119–125.

THOMPSON, E. A. and GEYER, C. J. (2005). Fuzzy p-values in latent variable problems. Technical Report 481, Dept. Statistics, Univ. Washington. Available at www.stat.washington.edu/www/research/reports/2005/tr481.pdf.

WHITTEMORE, A. and HALPERN, J. (1994). A class of tests for linkage using affected pedigree members. *Biometrics* **50** 118–127.