EXTENDING EDWARDS LIKELIHOOD RATIOS TO SIMPLE ONE SIDED HYPOTHESIS TESTS

DEWAYNE DERRYBERRY AND MILAN BIMALI

Abstract. With regard to the one sided hypothesis test, we propose a likelihood ratio that might be viewed as a Bayes/Non-Bayes compromise in the spirit of I. J. Good (1983). The influence of A. W. F. Edwards (1972) will also be apparent. Although we will develop some general ideas, most of our effort will focus on tests of a single unknown mean and the specific case of a sample from a normal population with unknown mean and known variance.

1. The Framework

Suppose we have a random sample from a normal distribution with known variance and unknown mean (NUK). Neyman and Pearson proposed an NUK test of the form $H_0 : \mu = \mu_1$ versus $H_1 : \mu = \mu_2$. Neyman and Pearson [2] framed the problem as that of making a decision about the two hypotheses, with the underlying notion of type I and type II error associated with the decision. In this framework, if the p-value is less than some preset level, called α, a decision will be made to adopt the second hypothesis, if the p-value is greater than α, the decision will be made to adopt the first hypothesis.

Fisher proposed a different way of thinking about hypotheses of the form $H_0 : \mu = \mu_0$ with no explicit alternative hypothesis given. In this context a p-value is also computed (as a two-sided p-value) and we claim to have evidence against the null hypothesis when the p-value is sufficiently small, and take a neutral stance if the p-value is large. In no case can there be evidence for the null hypothesis in this context. It is now well understood that the approaches of Neyman-Pearson and Fisher are completely different [2].

The most common test found in introductory textbooks is of the form of either $H_0 : \mu = \mu_0$ or versus $H_1 : \mu \neq \mu_0$ (the two forms of the formulation already hint at a possible conceptual confusion), and can be viewed as some hybrid of these two ideas. These tests are often presented with α levels and decisions (the Neyman-Pearson approach) or with the p-values and stated levels of evidence for the alternative hypothesis. Often, textbooks alternate between these two approaches without explanation, this can be verified with almost any introductory or intermediate level textbook.

2. A. W. F. Edwards Approach

Edwards [4] identified a way of thinking about simple tests of the form $H_1 : \mu = \mu_1$ versus $H_2 : \mu = \mu_2$ which are different from the approach of Neyman and Pearson. Edwards is a frequentist and does not, in general, believe in the doctrine of inverse probability, the use of Bayes Theorem to derive $Pr(H|\text{data})$ from $Pr(\text{data}|H)$ but does believe there are specific cases where inverse probability is possible [4, p. 44].

Furthermore Edwards, as a scientist, has little interest in statistical analysis as a decision making exercise. In fact, his contempt for the decision theory approach is apparent from the very first pages of his book. Decisions are often based on statistical analysis (clinical trials, process optimization, quality control, etc), but practicing scientists are often more interested in compiling and summarizing many studies, not in making some sort of immediate decision about each individual study Rosenbaum [9] and Ramsey and Schafer [8, pp. 48–49].

Where Neyman and Pearson were thinking of the comparison of the hypotheses as resulting in a decision about which hypothesis to act upon, Edwards was thinking more along the lines of the relative evidence for each hypothesis and proposed that likelihood ratios were a way of assessing the relative odds of the two hypothesis using Bayes Theorem (along with a prior belief each hypothesis was equally likely). He reasoned

$$\frac{Pr(H_1|data)}{Pr(H_2|data)} = \frac{L(data|H_1)Pr(H_1)}{L(data|H_2)Pr(H_2)} = \frac{L(data|H_1)}{L(data|H_2)}$$
$$= LR_{12}(\text{the likelihood ratio}) \quad -(1)$$

is a measure of the odds of the hypotheses, given the data. $(L(data|H_i), i = 1, 2$ is the likelihood of the data under specified hypothesis).

Although the argument involves inverse probability, it is among the least offensive to frequentists. Indeed, if all likelihood ratios had a simple interpretation as the relative odds of the two competing hypotheses, and all priors were so uncontroversial, there might not be a distinction between Bayesians and Frequentists in this special context.

Expression (1) would now be recognized as a Bayes factor [5, 6, 7]. However, the form is particularly pure, because we do not need to form an integral involving the prior, nor is the prior particularly subjective.

In particular Good [5, 6] identifies this specific ratio, without the simplifying assumption $Pr(H_1) = Pr(H_2)$ as a Bayes factor, as an odds ratio comparing the weight of evidence in favor of each hypothesis, and as something non-Bayesian could accept in the case of two simple hypotheses. Incidentally, despite the apparently similar intentions of Edwards and Good, neither author cited the other in our references, although two of the references are books and both discuss this ratio extensively in these books.

3. Some Useful Properties of Likelihood Ratio (NUK Case)

(1) $\text{LR}_{12} = 1$ when $\frac{\mu_1+\mu_2}{2} = \bar{x}$ and both hypotheses are equally likely.
(2) If $\text{LR}_{12} > 1$, then H_1 is more plausible than H_2 and vice-versa.
(3) If H_1 is true, as $n \to \infty$, $\text{LR}_{12} \to \infty$. If H_2 is true, as $n \to \infty$, $\text{LR}_{12} \to 0$.
(4) $Pr(H_1|\text{data}) = \frac{\text{LR}_{12}}{1+\text{LR}_{12}}$ and $Pr(H_2|\text{data}) = 1 - Pr(H_1|\text{data})$

4. Another Way of Thinking about Inverse Probability

For the non-Bayesian, another kind of evidence, related to the basic ratio, could be useful. DeGroot and Schervish [3, pp. 304–307] discuss a court case in which the amount of prior weight must be placed on a hypothesis, for the posterior probability to be reasonable. For example, we could consider the minimum value $Pr(H_1)$ such that

$$\frac{Pr(H_1|data)}{Pr(H_2|data)} = \frac{L(Data|H_1)Pr(H_1)}{L(data|H_2)(1 - Pr(H_2))} = 1.$$

Good [5, pp. 36–37] argues similarly for a value of inverse probabilities even when priors are not known. Here, as before, the fact that the prior is a fixed value, and not a continuous distribution, makes things more manageable and less controversial.

5. One Sided Tests

We accept the argument made by Edwards (and reinforced by Good), and wish to extend the idea to a new context. We also accept the following notions held by Edwards. Entities like Bayes factors or likelihood ratios are more informative that reject/don't reject decisions and that inverse probability arguments are sometimes possible, but a non-Bayesian needs a rationale for the plausibility of the prior in each case.

Consider an alternative formulation of the typical Neyman-Pearson test: $H_1 : \mu = \mu_1$ versus $H_2 : \mu = \mu_2$. Let $c = \frac{\mu_1+\mu_2}{2}$, then the test could be stated as: $H_1 : \mu = c + \Delta$ versus $H_2 : \mu = c - \Delta$ where $\Delta > 0$ and the likelihood ratio would be:

$$\frac{Pr(H_1|data)}{Pr(H_2|data)} = \frac{L(data|H_1)}{L(data|H_2)} = \text{LR}_{12}.$$

Specializing to NUK case we have $\text{LR}_{12} = exp\left(2\Delta n(c - \bar{x})/\sigma^2\right)$. The term $(c - \bar{x})$ determined whether LR_{12} is greater than, or less than 1. We now consider a variation of this situation. We still have the hypotheses: H_1 : $\mu = c + \Delta$ versus $H_2 : \mu = c - \Delta$, but suppose now that c is known but $\Delta > 0$ and is unknown (this is the matter in which introductory textbooks present such tests). This is similar to the test of the form $H_1 : \mu < c$ versus $H_2 : \mu > c$. We assume $Pr(H_1) = Pr(H_2) = 0.5$, as did Edwards. In many real problems the choice of c is obvious, we want to assume (due to fairness) that prior to the collection of the data we show no bias (a prior judicial belief that each hypothesis is equally likely). In this case the likelihood ratio becomes:

$$\frac{L(data|H_1)}{L(data|H_2)} = \frac{c - \hat{\Delta}}{c + \hat{\Delta}} = \mathrm{LR}^*.$$

Specializing the result to NUK, we have $\hat{\Delta} = |c - \bar{x}|$ and

$$\mathrm{LR}^* = exp\left(2\hat{\Delta}n(c - \bar{x})/\sigma^2\right),$$

this is the likelihood ratio for a Neyman-Pearson test in which μ_1 and μ_2 are unknown but the midpoint between them is known.

6. Some Useful Properties of Likelihood Ratio (NUK)

(1) $\mathrm{LR}^* = 1$ when $c = \bar{x}$ and both hypotheses are equally likely.
(2) If LR > 1, then $c > \bar{x}$ and H_1 is more plausible than H_2 and vice-versa.
(3) If H_1 is true, as $n \to \infty, \mathrm{LR}^* \to \infty$. If H_2 is true, as $n \to \infty$, $\mathrm{LR}^* \to 0$.
(4) We know $\bar{x} \to \mu$, so $\hat{\Delta} \to \Delta$, and $\sqrt[n]{\mathrm{LR}^*_{12}} \to \sqrt[n]{\mathrm{LR}_{12}}$. So that asymptotically

$$\begin{aligned} Pr^*(H_1|Data) &= (\mathrm{LR}^*_{12}/(1 + \mathrm{LR}^*_{12})) = (\mathrm{LR}_{12}/(1 + \mathrm{LR}_{12})) \\ &= Pr(H_1|Data). \end{aligned}$$

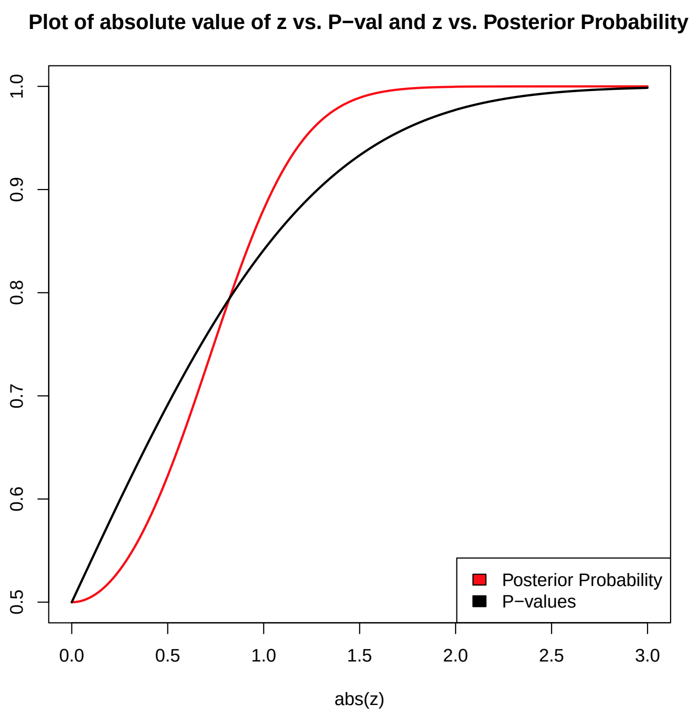
7. Bayesian Factors, Posterior Probabilities, and p-values

If any continuous prior is placed on Δ such that $\Delta \geq 0$ and each hypothesis is equally likely so that half the time $\mu = c - \Delta$, and half the time $\mu = c + \Delta$, could produce a Bayes factor. Our approach does not place a prior on Δ, but treats it as fixed, but unknown. We do, however, assume that half the time $\mu = c - \Delta$, and half the time $\mu = c + \Delta$.

The relationship between frequentist p-values and Bayesian posterior probabilities has been the topic of much discussion [1, p. 414]. In broad strokes, for two-sided tests p-values are generally much smaller than Bayesian posterior probabilities. On the other hand, for one sided tests p-values can be larger or smaller than posterior probabilities. It is a simple task to compute both p-values and posterior probabilities using the formulas found in this paper, because $Pr^*(H_1|Data)$ and the p-value are both functions of $z - score = (\bar{x} - c)\sqrt{n}/\sigma$.

TABLE 1. $H_1 : \mu < c$ and $H_2 : \mu > c$. A comparison of p-values and posterior prababilities

$z - score$	LR*	p-value	$Pr^*(H_1|Data)$
0.00	1.000	0.50	0.50
-0.45	1.500	0.674	0.60
-0.65	2.333	0.742	0.70
-0.83	4.000	0.797	0.80
-1.05	9.000	0.853	0.90
-1.21	19.00	0.887	0.95
-1.52	99.00	0.936	0.99
-1.62	199.0	0.947	0.995



Although not much can be said from Table 1 and Figure 1, as we know there is no particular expected relationship between one sided p-values and

Bayesian posterior probabilities, so there is no pattern to confirm, nevertheless the comparison is intrinsically interesting. The table and the plot (Figure 1) show that, as the z-score increases in magnitude, the posterior probability grows faster than the p-value.

8. Summary

We begin with the assumption of Edwards [4, Chapter 4] that the original derivation is sound, a position confirmed by Good [5, 6]. In that derivation Bayes Theorem, and a prior assumption both hypotheses are equally likely to be true, were combined to produce a likelihood ratio that can be understood as a Bayes factor and also the relative odds of the two hypotheses given the data.

By reformulating the Neyman-Pearson framework, we develop an Edwards-like approach for the case of one-sided tests. We would argue that our approach is a minimal extension of Edwards program and allows for the computation of inverse probabilities. We would argue that at least some non-Bayesians will accept our development of an inverse probability, one that does not put a prior on the unknown parameter. We further view this as an example of a Bayes/non-Bayes compromise, often advocated by Good.

9. Acknowledgments

The authors are grateful to the statistics referees of the Missouri Journal of Mathematical Sciences for their role in polishing the paper and strengthening its results.

References

[1] G. Casella and R. Berger, *Reconciling Bayesian and frequentist evidence in the one-sided testing problem*, Journal of the American Statistical Association, **82.397** (1987), 106–111.

[2] R. Christensen, *Testing, Fisher, Neyman, Pearson, and Bayes*, The American Statistician, **59.2** (2005), 121–126.

[3] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, 3rd ed., Addison-Wesley, 2002.

[4] A. W. F. Edwards, *Likelihood*, Cambridge University Press, Cambridge (expanded edition, 1992, Johns Hopkins University Press, Baltimore), 1972.

[5] I. J. Good, *Good Reasoning: The Foundations of Probability and Its Applications*, University of Minnesota Press, 1983.

[6] I. J. Good, *The Bayes/non-Bayes compromise: a brief review*, Journal of the American Statistical Association, **87.419** (1992), 576–606.

[7] R. E. Kass and A. E. Raftery, *Bayes factors*, Journal of the American Statistical Association, **90.430** (1995), 791.

[8] F. L. Ramsey and D. W. Schafer, *The Statistical Sleuth*, Brooks/Cole, 1953.

[9] P. R. Rosenbaum, *Observational Studies*, 2nd ed., Springer, New York, 2002.

MSC2010: 62A01

Keywords: Bayes factor, inverse probability, weight of evidence.

Appendix

Note as long as $\frac{\Delta^2 n}{\sigma^2}$ is constant,
$log(\mathrm{LR}_{12}) = -\frac{2\Delta n(\bar{x}-c)}{\sigma^2} = -\frac{2\Delta\sqrt{n}}{\sigma}\frac{\sqrt{n}(\bar{x}-\mu+\mu-c)}{\sigma} = -\frac{2\Delta\sqrt{n}}{\sigma}z + \frac{2\Delta^2 n}{\sigma^2}$ is a normal random variable with mean and variance:

$$E\left(log(\mathrm{LR}_{12})\right) = E\left(-\frac{2\Delta\sqrt{n}}{\sigma}z + \frac{2\Delta^2 n}{\sigma^2}\right) = \frac{2\Delta^2 n}{\sigma^2}$$

$$Var\left(log(\mathrm{LR}_{12})\right) = E\left(-\frac{2\Delta\sqrt{n}}{\sigma}z + \frac{2\Delta^2 n}{\sigma^2}\right) = \frac{4\Delta^2 n}{\sigma^2}$$

Now: $log(\mathrm{LR}^*_{12}) = \frac{2\hat{\Delta}n(\bar{x}-c)}{\sigma^2} = \frac{2n(\bar{x}-c)^2}{\sigma^2} = \frac{2n(\bar{x}-\mu+\mu-c)^2}{\sigma^2} = 2z^2 - \frac{4\Delta\sqrt{n}}{\sigma^2}z + \frac{2\Delta^2 n}{\sigma^2}$

The estimated $log(\mathrm{LR})$ is a linear combination of χ^2_1 and a standard normal plus a constant, where the normal dominated for large n. The mean and variance are:

$$E\left(log(LR^*_{12})\right) = E\left(2z^2 - \frac{4\Delta\sqrt{n}}{\sigma}z + \frac{2\Delta^2 n}{\sigma^2}\right) = 2 - 0 + \frac{2\Delta^2 n}{\sigma^2}$$

$$Var\left(log(LR^*_{12})\right) = Var\left(2z^2 - \frac{4\Delta\sqrt{n}}{\sigma}z + \frac{2\Delta^2 n}{\sigma^2}\right) = 4.Var(\chi^2_1) + \frac{16\Delta^2 n}{\sigma^2} = 8 + \frac{16\Delta^2 n}{\sigma^2},$$

z and z^2 are uncorrelated.

Department of Mathematics, Idaho State University, Pocatello, ID
E-mail address: derrdewa@isu.edu

Department of BioStatistics, University of Kansas - Medical Center, Kansas City, KS 66160
E-mail address: mbimali@kumc.edu