Prof. David Draper
Department of Statistics
University of California, Santa Cruz

STAT 131: Quiz 4 *[40 total points]*

Name: _____

- *(Fact 1)* As a broad generalization (which you can verify empirically), statisticians tend to have shy personalities more often than economists do — let's quantify this observation by assuming that 80% of statisticians are shy but the corresponding percentage among economists is only 15%.
- *(Fact 2)* Conferences on the topic of *econometrics* are almost exclusively attended by economists and statisticians, with the majority of participants being economists — let's approximately quantify this fact by assuming that 90% of the attendees are economists (and the rest statisticians).

Suppose that you (a physicist, say) go to an econometrics conference — you strike up a conversation with the first person you (haphazardly) meet, and find that this person is shy. The point of this problem is to show that the (conditional) probability p that you're talking to a statistician is only about 37%, which most people find surprisingly low, and to understand why this is the right answer. Let St = (person is statistician), E = (person is economist), and Sh = (person is shy).

(a) Identify the most important assumption needed in this problem to permit its solution to be probabilistic; explain briefly. *[5 points]*

(b) Using the St, E and Sh notation, express the three numbers (80%, 15%, 90%) above, and the probability we're solving for, in unconditional and conditional probability terms. *[5 points]*

(c) Briefly explain why calculating the desired probability is a good job for Bayes's Theorem. (*Hint:* Is there a true/false statement (proposition) here whose truth is unknown to you? Is there a proposition that plays the role of data?) *[5 points]*

(d) Briefly explain why the following expression is a correct use of Bayes's Theorem in odds form in this problem.

$$\begin{array}{ccccc} \left[\frac{P(St \mid Sh)}{P(E \mid Sh)}\right] & = & \left[\frac{P(St)}{P(E)}\right] & \cdot & \left[\frac{P(Sh \mid St)}{P(Sh \mid E)}\right] \\ (1) & = & (2) & \cdot & (3) \end{array}$$

(e) Here are three terms that are relevant to the quantities in part (c) above:

- (Prior odds ratio in favor of St over E)
- (Bayes factor in favor of St over E)
- (Posterior odds ratio in favor of St over E)

Match these three terms with the numbers $(1), (2), (3)$ in the second line of the equation in part (d). *[5 points]*

(f) Compute the three odds values in part (e), briefly explaining your reasoning, thereby demonstrating that the posterior odds ratio o in favor of St over E is $o = \frac{16}{27} \doteq 0.593$. *[5 points]*

(g) Use the expression $p = \frac{o}{1+o}$ to show that the desired probability in this problem — the conditional probability that you're talking to a statistician — is $p = \frac{16}{43} \doteq 0.372$. *[5 points]*

(h) Someone says, "That probability can't be right: 80% of statisticians are shy, versus 15% for economists, so your probability of talking to a statistician has to be over 50%." Briefly explain why this line of reasoning is wrong, and why p should indeed be less than 50%. *[5 points]*