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

STAT 206: Quiz 1 *[90 total points]*

Name: Kevin Guillen

Here is Your background information, translatable into $\mathcal{B}$, for this problem.

- *(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 (based on previous psychological studies) that 84% of statisticians are shy but the corresponding percentage among economists is only 10%.

- *(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 (based on data from previous conferences) that 95% of the attendees are economists (and the rest statisticians, except for a tiny proportion of people from other professions, which can be ignored).

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, given this data and the above background information, is only about 31%, 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 (in the form of a proposition B_1, one of the elements of $\mathcal{B}$) the most important assumption needed in this problem to permit its solution to be probabilistic; explain briefly. *[5 points]*

Solution. Our most important assumption is that we haphazardly meet a person. Since translating this requirement to math language our B_1 is taking a random sample of 1 from the population, which is all conference attendees. $\square$

(b) Using the St, E and Sh notation, express the three numbers (84%, 10%, 95%) above, and the probability we're solving for, in conditional probability terms, remembering to condition appropriately on $\mathcal{B}$. *[5 points]*

Solution.

$$84\% = P(Sh \mid St, \mathcal{B}) \tag{1}$$
$$10\% = P(Sh \mid E, \mathcal{B}) \tag{2}$$
$$95\% = P(E \mid \mathcal{B}) \tag{3}$$

we also obtain the following,

$$5\% = P(St \mid \mathcal{B}) \tag{4}$$

$\square$

(c) Briefly explain why calculating the desired probability is a good job for Bayes's Theorem. *[5 points]*

Solution. Well, we are trying to calculate $P(St \mid Sh, \mathcal{B})$ and we see that probability (1) from above is just the reverse ordering of conditioning of what we are trying to solve for, which is perfect for Bayes' Theorem.

$$P(St \mid Sh, \mathcal{B}) = \frac{P(St \mid \mathcal{B})P(Sh \mid St, \mathcal{B})}{P(Sh \mid \mathcal{B})}$$

$\square$

Table 1: *2 by 2 table cross-tabulating truth (statistician, economist) against data (shy, not shy) for the people at the conference, assuming a total number of attendees of 1,000.*

		Truth		
		Statistician	Economist	Total
Data	Shy	42	95	137
	Not Shy	8	855	863
	Total	50	950	1,000

The goal in the rest of the problem is for you to use all three of the methods developed in class — the 2 by 2 table cross-tabulating truth against data, Bayes's Theorem in odds ratio form, and calculating the denominator using the *Law of Total Probability*, by partitioning over the unknown truth — to compute $P(St \mid Sh, \mathcal{B})$, the posterior probability that the haphazard person is a statistician given that this person is shy (and given $\mathcal{B}$).

(d) Use the three numerical facts (84%, 10%, 95%) given at the beginning of the quiz to fill in all 8 of the entries marked '—' in Table 1, taking the total number of attendees at the conference to be 1,000 (*Hint:* All of these numbers are integers), thereby showing that $P(St \mid Sh, \mathcal{B}) = \frac{42}{137} \doteq 30.7\%$; show your work *[20 points]*.

Solution. First we can obtain the total number of economists and statisticians using (3) and (4) from part (b),

$$|E| = P(E \mid \mathcal{B}) \cdot 1000 = 0.95 \cdot 1000 = 950$$
$$|St| = P(St \mid \mathcal{B}) \cdot 1000 = 0.05 \cdot 1000 = 50.$$

Now with (1) and (2) we can get how many of these statisticians and economists are shy,

$$|Sh, St| = P(Sh \mid St, \mathcal{B}) \cdot |St| = .84 \cdot 50 = 42$$
$$|Sh, E| = P(E \mid E, \mathcal{B}) \cdot |E| = .10 \cdot 950 = 95$$

now knowing the total of economists and statisticians alongside of many of each are shy we can obtain how many of them are not shy,

$$|\neg Sh, St| = |St| - |Sh, St| = 50 - 42 = 8$$
$$|\neg Sh, E| = |E| - |Sh, E| = 950 - 95 = 855.$$

Which means given that the person we haphazardly meet at this convention is shy, the probability that they are a statistician is,

$$P(St \mid Sh, \mathcal{B}) = \frac{42}{137} = 30.7\%$$

□

(e) Briefly explain why the following expression is a correct use of Bayes's Theorem on the odds ratio scale in this problem. *[5 points]*

$$\begin{array}{ccccc} \left[\dfrac{P(St \,|\, Sh, \mathcal{B})}{P(E \,|\, Sh, \mathcal{B})} \right] & = & \left[\dfrac{P(St \,|\, \mathcal{B})}{P(E \,|\, \mathcal{B})} \right] & \cdot & \left[\dfrac{P(Sh \,|\, St, \mathcal{B})}{P(Sh \,|\, E, \mathcal{B})} \right] \\ (1) & = & (2) & \cdot & (3) \end{array}$$

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

- (Prior odds ratio in favor of St over E, given $\mathcal{B}$)
- (Bayes factor in favor of St over E, given $\mathcal{B}$)
- (Posterior odds ratio in favor of St over E, given $\mathcal{B}$)

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

(g) Compute the three ratios in part (e), briefly explaining your reasoning, thereby demonstrating that the posterior odds ratio o in favor of St over E (given $\mathcal{B}$) is $o = \frac{42}{95} \doteq 0.442$. *[15 points]*

(h) 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 (given $\mathcal{B}$) — is $p = \frac{42}{137} \doteq 0.307$. *[5 points]*

(i) Briefly explain why the following expression is a correct use of Bayes's Theorem on the probability scale in this problem. *[5 points]*

$$P(St \,|\, Sh, \, \mathcal{B}) = \frac{P(St \,|\, \mathcal{B}) \, P(Sh \,|\, St, \, \mathcal{B})}{P(Sh \,|\, \mathcal{B})} \,. \tag{5}$$

(j) Notice as usual that you know both of the numerator probabilities in equation (6) but you don't (yet) know the denominator $P(Sh \,|\, \mathcal{B})$. Use the *Law of Total Probability*, partitioning over the unknown truth, to show that

$$P(Sh \,|\, \mathcal{B}) = \frac{137}{1000} = 0.137\,, \tag{6}$$

and use this to show that

$$P(St \,|\, Sh,\, \mathcal{B}) = \frac{(0.05)(0.84)}{0.137} = \frac{42}{137} \doteq 0.307\,. \tag{7}$$

[15 points]

(k) Someone says, "That 30.7% probability can't be right: 84% of statisticians are shy, versus 10% for economists, so your probability p 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]*