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

STAT 131: Quiz 3 *[15 total points]*

Name: ____

Bayes's Theorem is the only known approach to learning from data that satisfies two important properties: (a) it's logically internally consistent (meaning that it cannot produce contradictory conclusions such as {An unknown quantity θ of interest to me cannot be negative, but Bayes's Theorem says that my best estimate of θ is -2.3}, and (b) it combines information external and internal to your dataset in such a way that no extraneous information is inadvertently smuggled into your answer. However, it's possible to use Bayes's Theorem in a way that defeats its ability to help you learn from the world around you. The result we'll explore below in this quiz that illustrates this was called *Cromwell's Rule* by the great British Bayesian statistician Dennis Lindley (1923–2013).

Let U be a true-false proposition whose truth status is Unknown to you, and let D be another true-false proposition (representing Data) whose truth status is known to you; an example would be $U =$ (person P really is infected with COVID–19) and $D =$ (this PCR screening test *says* that person P is infected with COVID–19) (note that U and D are not the same; the PCR test could be wrong). Recall that Bayes's Theorem in this situation says, assuming that $P(D) > 0$, that

$$P(U \mid D) = \frac{P(U)\, P(D \mid U)}{P(D)}\,, \tag{1}$$

in which $P(U)$ is your *prior* information about the truth of U (in the example above, this would be the prevalence of COVID–19 among people similar to person P in all relevant ways).

(a) Show that if you assume that $P(U) = 0$, then you would have to conclude that $P(U \mid D) = 0$, no matter how the data information D came out *[5 points]*.

(b) Show that if you assume that $P(U) = 1$, then you would have to conclude that $P(U \mid D) = 1$, no matter how the data information D came out *[5 points]*.

(c) Briefly explain in what sense your results in (a) and (b) imply that

> *Putting prior probability 0 or 1 on anything renders Bayes's Theorem incapable of helping you learn from data.*

What practical conclusion should we draw about the assignment of prior probabilities? Explain briefly *[5 points]*.