Astrostatistics solutions

Roberto Trotta, Imperial College London

November 14, 2011

1 Probabilistic reasoning

1. A batch of chemistry undergraduates are screened for a dangerous medical condition [...]

   Answer

   Let $BB = 1$ denote the proposition that your friend has the virus, and $BB = 0$ that she does not. We use $+$ $(-)$ to denote the test returning a positive (negative) result.

   We know from the reliability of the test that

   $$P(+|BB = 1) = 0.95 \quad (1)$$
   $$P(-|BB = 0) = 0.95. \quad (2)$$

   Given that 1% of the population has the virus, the probability of being one of them (before taking the test) is $P(BB = 1) = 0.01$, while $P(BB = 0) = 0.99$.

   The probability of your friend having the virus after she has tested positive is thus

   $$P(BB = 1|+) = \frac{P(+|BB = 1)P(BB = 1)}{P(+)}. \quad (3)$$

   We can compute the denominator as follows:

   $$P(+) = P(+|BB = 1)P(BB = 1) + P(+|BB = 0)P(BB = 0) = 0.95 \cdot 0.01 + 0.05 \cdot 0.99 = 0.059. \quad (4)$$

   Therefore the probability that your friend has the virus is much less than 95%, namely

   $$P(BB = 1|+) = \frac{0.95 \cdot 0.01}{0.059} = 0.16 = 16\%. \quad (5)$$

2. A pan contains 10 ravioli, of which 9 are filled with pesto and one with ricotta [...]

   Answer

   Let us denote by p a pesto raviolo, by r a raviolo filled with ricotta. Then the joint probability of drawing two pesto ravioli is

   $$P(1st = p, 2nd = p) = P(2nd = p|1st = p)P(1st = p). \quad (6)$$

   The probability that the first raviolo is filled with pesto is $P(1st = p) = 10/11$, while for the second (conditional) draw $P(2nd = p|1st = p) = 9/10$. So the combined probability is $P(1st = p, 2nd = p) = 9/11$.

In the second case, for the raviolo in your plate (yr) you have a prior $P(yr = p) = P(yr = r) = 1/2$. After you have picked the raviolo and found it to be pesto $(p = 1)$ your posterior probability for the remaining raviolo in your plate is:

$$
\begin{aligned}
P(yr = p|p = 1) &= \frac{P(p = 1|yr = p)P(yr = p)}{P(p = 1|yr = p)P(yr = p) + P(p = 1|yr = r)P(yr = r)} \\
&= \frac{1}{1 + \frac{P(p=1|yr=r)P(yr=r)}{P(p=1|yr=p)P(yr=p)}} = \frac{1}{1 + \frac{\frac{1}{2}\cdot\frac{1}{2}}{1\cdot\frac{1}{2}}} = 2/3.
\end{aligned}
\tag{7}
$$

3. In a TV debate, a politician named Barack affirms that climate change is caused by human activities [...]

   Answer

   Let CCH denote the proposition "climate change is caused by humans". Let B_T denote the statement "Barack tells the truth", B_L denote the statement "Barack lies" and similarly for S_T and S_L for Silvio. Under your prior, $P(B_T) = 2/3$, $P(B_L) = 1/3$, $P(S_T) = 1/4$ and $P(S_L) = 3/4$. Let "SCCH“ denote the statement "Silvio says CCH is true". Then using Bayes theorem:

$$
P(CCH|SCCH) = \frac{P(SCCH|CCH)P(CCH)}{P(SCCH)}. \tag{8}
$$

   In the above equation, $P(CCH) = P(B_T) = 2/3$, as you don't know anything else about statement CCH except what you heard from Barack, whom you trust to be truthful with probabilty $P(B_T)$. Also, $P(SCCH|CCH) = P(S_T) = 1/4$, for Silvio will say that statement CCH is true (if this is indeed the case) with probability $P(S_T)$. It remains to compute

$$
\begin{aligned}
P(SCCH) &= P(SCCH|CCH)P(CCH) + P(SCCH|\text{not } CCH)P(\text{not } CCH) \\
&= P(S_T)P(B_T) + P(S_L)P(B_L).
\end{aligned}
\tag{9}
$$

   So the posterior probability for CCH to be true after you have heard both politicians is

$$
\begin{aligned}
P(CCH|SCCH) &= \frac{P(S_T)P(B_T)}{P(S_T)P(B_T) + P(S_L)P(B_L)} \\
&= \frac{1}{1 + \frac{P(S_L)}{P(S_T)}\frac{P(B_L)}{P(B_T)}} = \frac{1}{1 + \frac{3}{2}} = 2/5 = 40\%.
\end{aligned}
\tag{10}
$$

   From the above, it is clear that your posterior belief in the proposition is influenced by the ratio of your relative trust in the source of the information (in this case, the politicians). In the absence of all information about climate change, one might have an indifference prior $P(CCH) = 1/2$. But after having acquired information about CCH from two sources which are considered relatively unreliable, one's posterior belief in CCH actually decreases, despite the fact that both politicians agreed in stating the reality of human-made climate change.

   This explains why people of different political orientation (who therefore hold divergent priors on the truthfulness of e.g. politicians, media, cable TV's etc) might find their respective positions radicalized after having been exposed to the same information.

4. You are playing poker [...]

   Answer

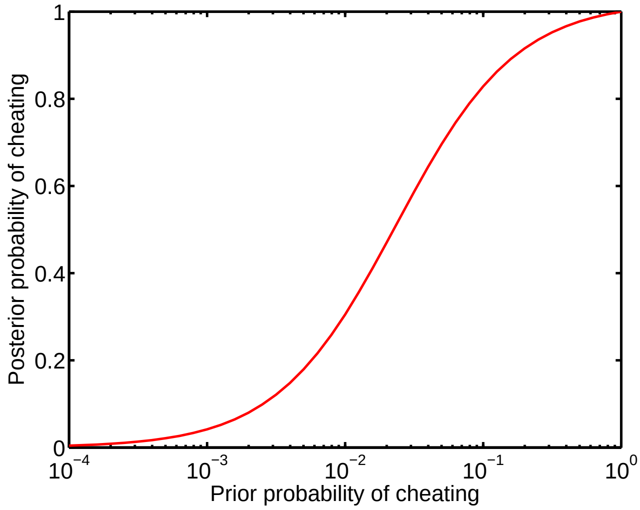

Figure 1: Posterior probability of cheating as a function of your prior $P(OC)$.

Let's denote by OC the probability that your opponent is cheating, and by ToK the event of the dealer getting a Three of a Kind. Then by Bayes theorem:

$$P(OC|ToK) = \frac{P(ToK|OC)P(OC)}{P(ToK|OC)P(OC) + P(ToK|\text{not } OC)P(\text{not } OC)} = \frac{1}{1 + \frac{P(ToK|\text{not } OC)(1-P(OC))}{P(ToK|OC)P(OC)}} \quad (11)$$

From the problem, we know that $P(ToK|\text{not } OC) \approx 1/46 = 0.021$. The probability of the dealer giving himself a Three of a Kind if he is cheating can be estimated as $P(ToK|OC) = 0.95$ (allowing for a 5% probability that even if the guy is cheating he might get the trick wrong - conceivably this probability might be different for different opponents, depending on what you think about their cheating skills). So we have that

$$P(OC|ToK) = \frac{1}{1 + 0.023\frac{1-P(OC)}{P(OC)}}. \quad (12)$$

This is plotted as a function of your prior probability that your opponent might be a cheater, $P(OC)$, in Fig. 1.

If you opponent is St Augustin, then perhaps $P(OC) = 10^{-3}$ (or less), and your posterior is $P(\text{St Augustin cheating}|T$ 0.042; against your brother, say your prior is $P(OC) = 0.5$, then $P(\text{your brother is cheating}|ToK) =$ 0.978; finally, against Al Capone, with $P(OC) = 0.9$ you get $P(\text{Al Capone cheating}|ToK) = 0.998$.

5. A body has been found on the Baltimore West Side [...]

   Answer

   Let us denote by $od = 1$ the statement "Mr Dunlop died because of drugs overdose"; by $Dd = 1$ the statement "Mr Dunlop is dead" and by $u = 1$ the statement "Mr Dunlop used drugs".

   We are looking for the posterior probability that Fuzzy Dunlop died of overdose, given that he was a drug addict ($u = 1$) and that he is dead ($Dd = 1$):

$$P(od = 1|Dd = 1, u = 1) = \frac{P(u = 1|od = 1, Dd = 1)P(od = 1|Dd = 1)}{P(u = 1|od = 1, Dd = 1)P(od = 1|Dd = 1) + P(u = 1|od = 0, Dd = 1)P(od = 0|Dd = 1)} \quad (13)$$

   From the problem, we have that the probability of being a drug user and having been murdered (assuming that people only die of either overdose or murder in Baltimore) is $P(u = 1|od = 0, Dd = 1) = 0.3$.

Also, the probability of the person having died of overdose (given that we have the body) is 50%, hence $P(od = 1|Dd = 1) = 50\%$ so $P(od = 0|Dd = 1) = 50\%$.

Finally, we need to estimate the probability that Mr Dunlop was a drug user, given that he died of overdose, $P(u = 1|od = 1, Dd = 1)$. It seems highly unlikely that somebody would die of overdose the first time they try drugs, so perhaps we can assign $P(u = 1|od = 1, Dd = 1) = 0.9$.

So we have that

$$P(od = 1|Dd = 1, u = 1) = \frac{1}{1 + \frac{P(u=1|od=0,Dd=1)P(od=0|Dd=1)}{P(u=1|od=1,Dd=1)P(od=1|Dd=1)}} = \frac{1}{1 + 3/9} = 75\%. \tag{14}$$

How sensitive is this conclusion to our estimate for $P(u = 1|od = 1, Dd = 1)$? Changing this to $P(u = 1|od = 1, Dd = 1) = 0.5$ (we are agnostic as to whether a drug overdose is more likely for usual drugs consumers or for novices) gives a posterior $P(od = 1|Dd = 1, u = 1) = 62\%$, while increasing it to $P(u = 1|od = 1, Dd = 1) = 0.99$ (most people overdosing are drugs users) gives $P(od = 1|Dd = 1, u = 1) = 77\%$.