Conditional Probability, Bayes' Theorem, and Unintuitive Results Solutions

1. Repeat to the breast cancer example, but now suppose that the prevalence was instead 0.1%. Do the same for a prevalence of 10%. Why can we always multiply the prior odds by the Bayes factor to get the posterior odds, but can not always do the same with the prior probability?

   Solution. For a prevalence of 0.1% the probability of having cancer given a positive test is about $\boxed{1\%}$ (odds $10 : 999$). For a prevalence of 10% that probability goes to around $\boxed{53\%}$ (odds $10 : 9$). The prior probability of having a disease $P(D)$ does not get updated via multiplying by the Bayes factor, but rather with the different likelihood ratio $\frac{P(+|D)}{P(+)}$. The methods using odds can be much more efficient, as we avoid needing to calculate $P(+)$.

2. An individual has been described by a neighbor as follows: "Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." Is Steve more likely to be a librarian or a farmer?

   Solution. Most people answer that Steve is more likely to be a librarian than a farmer. This is surely because Steve resembles a librarian more than a farmer, and associative memory quickly creates a picture of Steve in our minds that is very librarian-like. However, this is another example where conditionals can be erroneously reversed. There are five times as many farmers as librarians in the United States, and the ratio of male farmers to male librarians is even higher. With these priors in mind, it is clear that $\boxed{\text{Steve is more likely to be a farmer.}}$ That is, $P(\text{description} \mid \text{librarian})$ is high, but $P(\text{librarian} \mid \text{description})$ is relatively low.

3. Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in various protests. Rank the following scenarios from highest (1) to lowest (5) in probability. Linda is:

   - ___ an elementary school teacher
   - ___ active in the feminist movement
   - ___ a bank teller
   - ___ an insurance salesperson
   - ___ a bank teller also active in the feminist movement

   Solution. The main "gotcha" to watch out for here is to not put "bank teller and feminist" above "bank teller" or "feminist" individually. After all, this would violate a basic tenet of probability that $P(A \cap B) \leq P(A)$. Other than that, one can make reasonable arguments for placing the other options, though I would drift towards 2, 1, 3, 4, 5 (from top to bottom).

4. A couple has two children, the older of which is a boy. What is the probability that they have two boys?

   Solution. Let A be the event that both children are boys and B the event that the oldest child is a boy. We are interested in $P(A \mid B)$.

$$P(A \mid B) = P(A)\frac{P(B \mid A)}{P(B)} = \left(\frac{1}{4}\right)\left(\frac{1}{1/2}\right) = \boxed{\frac{1}{2}.}$$

5. A couple has two children, at least one of which is a boy. What is the probability that they have two boys? Is this the same as Problem 4?

   Solution. Let A be the event that both children are boys and C the event that at least one child is a boy. We are interested in $P(A \mid C)$. If the oldest child is a boy, there is at least one boy, but if there is at least one boy, that does not guarantee the oldest child is a boy. Thus, $B \neq C$. Note that we have $P(C) = 1 - P(\text{both children are girls}) = 1 - \frac{1}{4} = \frac{3}{4}$.

$$P(A \mid B) = P(A)\frac{P(C \mid A)}{P(C)} = \left(\frac{1}{4}\right)\left(\frac{1}{3/4}\right) = \boxed{\frac{1}{3}}.$$

6. A couple has two children, at least one of which is a boy, and that boy was born on a Tuesday. What is the probability that they have two boys? Is this the same as Problem 5?

   Solution. Let A be the event that both children are boys and D the event that at least one child is a boy born on Tuesday. We are interested in $P(A \mid D)$. We have $D \subseteq C$ but $C \nsubseteq D$, so $C \neq D$. Also note that $P(D \mid A)$ is not 1 this time. Let's assume that the probability of being born on a particular day of the week is $1/7$ and is independent of whether the child is a boy or a girl. Given that there are 7 days of the week, there are $7^2 = 49$ possible combinations for the days of the week the two boys were born on, and $(6)(1)+(1)(6)+(1)(1) = 13$ of these have at least one boy who was born on a Tuesday, so $P(D \mid A) = \frac{13}{49}$. To find $P(D)$, note that there are $(7+7)^2 = 14^2 = 196$ possible ways to select both the genders and the days of the week the two children were born on. Of these, there are $(6+7)^2 = 13^2 = 169$ ways which do not have a boy born on Tuesday. We have $196 - 169 = 27$, so $P(D) = \frac{27}{196}$. Thus,

$$P(A \mid D) = P(A)\frac{P(D \mid A)}{P(D)} = \left(\frac{1}{4}\right)\left(\frac{13/49}{27/196}\right) = \boxed{\frac{13}{27}}.$$

7. Balls numbered 1 through 20 are placed in a bag. Three balls are drawn out of the bag without replacement. What is the probability that all the balls have odd numbers on them?

   Solution (1). The probability that the first ball is odd is $1/2$. For the second ball, given that the first ball was odd, there are only 9 odd numbered balls that could be drawn from a total of 19 balls, so the probability is $9/19$. For the third ball, since the first two are both odd, there are 8 odd numbered balls that could be drawn from a total of 18 remaining balls, so the probability is $8/18$. Thus, the answer is

$$\left(\frac{1}{2}\right)\left(\frac{9}{19}\right)\left(\frac{8}{18}\right) = \boxed{\frac{2}{19}}.$$

   Solution (2). There are 10 odd numbers from 1 to 20 and we are selecting 3 numbers out of those 20 without replacement, so we can use the hypergeometric distribution and calculate

$$\frac{\binom{10}{3}\binom{10}{0}}{\binom{20}{3}} = \boxed{\frac{2}{19}}.$$

8. Zeb's coin box contains 8 fair, standard coins (heads and tails) and 1 coin which has heads on both sides. He selects a coin randomly and flips it 4 times, getting all heads. If he flips this coin again, what is the probability it will be heads?

   Solution (1). Let F and H_4 be the events of having the fair coin and flipping 4 heads, respectively, and denote the event of having the unfair coin as F^C. Via Bayes' Theorem,

$$\begin{aligned} P(F \mid H_4) = \frac{P(H_4 \mid F)P(F)}{P(H_4)} &= \frac{P(H_4 \mid F)P(F)}{P(H_4 \mid F)P(F) + P(H_4 \mid F^C)P(F^C)} \\ &= \frac{(1/16)(8/9)}{(1/16)(8/9) + (1)(1/9)} = \frac{1}{3}. \end{aligned}$$

   Thus, 1/3 is our new prior for the coin being fair. Let H be the event that the next flip is heads. Since there is a 1/3 probability of having chosen the fair coin given that 4 heads were flipped, the probability that the next flip is heads is

$$P(H) = \underbrace{\left(\frac{1}{3}\right)}_{\text{new prior}} \overbrace{\left(\frac{1}{2}\right)}^{P(H|F)} + \left(\frac{2}{3}\right) \underbrace{(1)}_{P(H|F^C)} = \boxed{\frac{5}{6}.}$$

   Solution (2).

$$P(\text{first 5 are heads} \mid \text{first 4 are heads}) = \frac{P(HHHHH)}{P(HHHH)} = \frac{\left(\frac{1}{9}\right)(1) + \left(\frac{8}{9}\right)\left(\frac{1}{2}\right)^5}{\left(\frac{1}{9}\right)(1) + \left(\frac{8}{9}\right)\left(\frac{1}{2}\right)^4} = \boxed{\frac{5}{6}.}$$

   Solution (3). For the first four coin tosses there are (9)(16) possible outcomes (9 for the choice of coin and 16 for the outcome of 4 tosses). We can ignore most of these, as only 8 of the outcomes will produce four heads from a fair coin and 16 outcomes will produce four heads from the fake coin. Each of these $8 + 16 = 24$ outcomes is equally likely and hence the probability that Zeb has picked a fair coin is only $8/24 = 1/3$. Now it is simply a matter of calculating the probability in the same manner as *Solution* (1).

9. There are 10 boxes containing blue and red balls. The number of blue balls in the n^{th} box is given by $B(n) = 2^n$. The number of red balls in the n^{th} box is given by $R(n) = 1024 - B(n)$. A box is picked at random, and a ball is chosen randomly from that box. If the ball is blue, what is the probability that the 10^{th} box was picked?

   Solution (1). Let A be the event that the 10^{th} box was picked and B the event that a blue ball was chosen. Using Bayes' theorem,

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{(1)(1/10)}{(1/10)\left(\frac{2^1+2^2+\ldots+2^{10}}{1024}\right)} = \frac{1024}{2^1 + 2^2 + \ldots + 2^{10}} = \frac{1024}{2046} = \boxed{\frac{512}{1023}.}$$

   Solution (2). Because each box has equal numbers of balls (1024), we can skip some calculation. The desired probability is

$$P(A \mid B) = \frac{B(10)}{B(1) + B(2) + B(3) + \ldots + B(10)} = \frac{1024}{2 + 4 + 8 + \ldots + 1024} = \frac{1024}{2046} = \boxed{\frac{512}{1023}.}$$

10. Suppose one has the training dataset given below:

Shape	Color	Size	Class
circle	blue	large	$+$
circle	red	medium	$-$
circle	red	large	$-$
square	blue	small	$-$
square	red	small	$-$
square	red	medium	$+$
square	blue	medium	$+$
square	blue	large	$-$
triangle	red	small	$+$
triangle	red	large	$+$
triangle	blue	medium	$+$

In our training set, our classes are not equiprobable in frequency. We have $P(+) = 6/11$ and $P(-) = 5/11$. We use this as our prior probability distribution. Classify the following sample as either $+$ or $-$ using MAP estimation and the naive Bayes assumption:

Shape	Color	Size	Class
circle	red	small	?

That is, determine which posterior is greater out of $+$ or $-$ for the above sample.

Solution. For $+$, we have $P(\text{circle} \mid +) = 1/6$, $P(\text{red} \mid +) = 1/2$, and $P(\text{small} \mid +) = 1/6$. We then calculate

$$\text{posterior numerator } (+) = \underbrace{\left(\frac{6}{11}\right)}_{P(+)}(\text{product of the above three}) = \frac{1}{132}.$$

Similarly for $-$, we have $P(\text{circle} \mid -) = 2/5$, $P(\text{red} \mid -) = 3/5$, and $P(\text{small} \mid -) = 2/5$. We then calculate

$$\text{posterior numerator } (-) = \underbrace{\left(\frac{5}{11}\right)}_{P(-)}(\text{product of the above three}) = \frac{12}{275}.$$

Since $\frac{12}{275} > \frac{1}{132}$, the $-$ class has a higher posterior, so we predict the sample is $\boxed{-.}$

★ 11. Suppose one has the training dataset given below:

Person	height (feet)	weight (lbs)	foot size (inches)
male	6	180	12
male	5.92 (5'11")	190	11
male	5.58 (5'7")	170	12
male	5.92 (5'11")	165	10
female	5	100	6
female	5.5 (5'6")	150	8
female	5.42 (5'5")	130	7
female	5.75 (5'9")	150	9

Let's say we have equiprobable classes, so $P(\text{male}) = P(\text{female}) = 0.5$. This prior probability distribution might be based on our knowledge of frequencies in the larger population, or on frequency in the training set. Assume feature distributions conditioned on the classes are Gaussian (that is, $P(\text{height} \mid \text{male})$ follows a normal distribution, and similarly for other feature-class pairs). Classify the following sample as either male or female using MAP estimation and the naive Bayes assumption:

Person	height (feet)	weight (lbs)	foot size (inches)
?	6	130	8

That is, determine which posterior is greater out of male or female for the above sample.

Solution. For the classification as male the posterior is given by

$$\text{posterior(male)} = \frac{P(\text{male})\, p(\text{height} \mid \text{male})\, p(\text{weight} \mid \text{male})\, p(\text{foot size} \mid \text{male})}{\text{evidence}}.$$

For the classification as female the posterior is given by

$$\text{posterior(female)} = \frac{P(\text{female})\, p(\text{height} \mid \text{female})\, p(\text{weight} \mid \text{female})\, p(\text{foot size} \mid \text{female})}{\text{evidence}}.$$

The evidence (also termed normalizing constant) may be calculated:

$$\begin{aligned}\text{evidence} = {} & P(\text{male})\, p(\text{height} \mid \text{male})\, p(\text{weight} \mid \text{male})\, p(\text{foot size} \mid \text{male}) \\ & + P(\text{female})\, p(\text{height} \mid \text{female})\, p(\text{weight} \mid \text{female})\, p(\text{foot size} \mid \text{female}).\end{aligned}$$

We have,

$$\begin{aligned} p(\text{height} \mid \text{male}) &= N(\mu_{hm}, \sigma_{hm}) \\ p(\text{weight} \mid \text{male}) &= N(\mu_{wm}, \sigma_{wm}) \\ p(\text{foot size} \mid \text{male}) &= N(\mu_{sm}, \sigma_{sm}) \\ &\vdots \end{aligned}$$

where $N(\mu, \sigma)$ is the Gaussian probability density function with mean μ and standard deviation σ. Given variances are unbiased sample variances and performing maximum likelihood estimation of the mean (which is also unbiased), we have

	height mean	height variance	weight mean	weight variance	foot size mean	foot size variance
male	$\mu_{hm} = 5.855$	$\sigma^2_{hm} = 0.0350$	$\mu_{wm} = 176.25$	$\sigma^2_{wm} = 122.9$	$\mu_{sm} = 11.25$	$\sigma^2_{sm} = 0.9167$
female	$\mu_{hf} = 5.418$	$\sigma^2_{hf} = 0.0972$	$\mu_{wf} = 132.5$	$\sigma^2_{wf} = 558.3$	$\mu_{sf} = 7.5$	$\sigma^2_{sf} = 1.667$

Given the sample, the evidence is a constant and thus scales both posteriors equally. It therefore does not affect classification and can be ignored. We now determine the probability density for the sex of the sample:

$$p(\text{height} = 6 \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma_{hm}^2}} e^{\frac{-(6-\mu_{hm})^2}{2\sigma_{hm}^2}} = 1.5789,$$

where $\mu_{hm} = 5.855$ and $\sigma_{hm}^2 = 3.5033 \cdot 10^{-2}$ are the parameters of normal distribution which have been previously determined from the training set. Note that since height, weight, and foot size are continuous variables, we use the lowercase p to represent density as opposed to the uppercase P representing probability. Thus, a value greater than 1 is fine here since it is a probability density rather than a probability. We similarly calculate

$$p(\text{weight} = 130 \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma_{wm}^2}} e^{\frac{-(130-\mu_{wm})^2}{2\sigma_{wm}^2}} = 5.9881 \cdot 10^{-6},$$

$$p(\text{foot size} = 8 \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma_{sm}^2}} e^{\frac{-(8-\mu_{sm})^2}{2\sigma_{sm}^2}} = 1.3112 \cdot 10^{-3},$$

$$\text{posterior numerator (male)} = \underbrace{(0.5)}_{P(\text{male})} (\text{product of the above three}) = 6.1984 \cdot 10^{-9}.$$

For females, we calculate in the same way $p(\text{height} = 6 \mid \text{female}) = 2.2346 \cdot 10^{-1}$, $p(\text{weight} = 130 \mid \text{female}) = 1.6789 \cdot 10^{-2}$, and $p(\text{foot size} = 8 \mid \text{female}) = 2.8669 \cdot 10^{-1}$. Thus, we have

$$\text{posterior numerator (female)} = \underbrace{(0.5)}_{P(\text{female})} (\text{product of the above three}) = 5.3778 \cdot 10^{-4}.$$

Since the posterior numerator is greater in the female case, we predict the sample is $\boxed{\text{female.}}$