University of Chicago

Business 41000: Business Statistics

Collection of Bayes questions

Problem B. (20 points)
Assume that the probability is 0.95 that a jury selected to try a criminal case will arrive at the correct verdict whether innocent or guilty. Further, suppose that the 80% of people brought to trial are in fact guilty.

1. Given that the jury finds a defendant innocent what's the probability that they are in fact innocent?
2. Given that the jury finds a defendant guilty what's the probability that they are in fact guilty?
3. Do these probabilities sum to one?

Let I denote the event of innocence and let VI denote the event that the jury proclaims an innocent verdict.

1. We want $P(I|VI)$ which is given by Bayes' rule

$$P(I|VI) = \frac{P(VI|I)P(I)}{P(VI)}$$

   where, by the law of total probability $P(VI) = P(VI|I)P(I) + P(VI|\bar{I})P(\bar{I})$ Hence $P(VI) = 0.95 \cdot 0.2 + 0.05 \cdot 0.8 = 0.23$
   Hence $P(I|VI) = \frac{0.95 \cdot 0.2}{0.23} = 0.83$

2. Similarly

$$P(G|VG) = \frac{P(VG|G)P(G)}{P(VG)} = \frac{P(VG|G)P(G)}{1 - P(VI)} = \frac{0.95 \cdot 0.8}{1 - 0.23} = 0.99$$

3. No. $P(I|VI) + P(G|VG) \neq 1$ because they are conditioned on different events.

Problem C. (20 points)
An oil company has purchased an option on land in Alaska. Preliminary geologic studies have assigned the following probabilities of finding oil

$$P(\text{high quality oil}) = 0.50 \quad P(\text{medium quality oil}) = 0.20 \quad P(\text{no oil}) = 0.30$$

After 200 feet of drilling on the first well, a soil test is taken. The probabilities of finding the particular type of soil identified by the test are as follows:

$$P(\text{soil} \mid \text{high quality oil}) = 0.20 \quad P(\text{soil} \mid \text{medium quality oil}) = 0.80 \quad P(\text{soil} \mid \text{no oil}) = 0.20$$

- What are the revised probabilities of finding the three different types of oil?
- How should the firm interpret the soil test?

1. *Using Bayes' Theorem, we get*

$$P(\text{high quality oil} \mid \text{soil}) = \frac{P(\text{soil} \mid \text{high quality oil})P(\text{high quality oil})}{P(\text{soil})}$$

Using the law of total probability

$$P(\text{soil}) = 0.5 \times 0.2 + 0.2 \times 0.8 + 0.3 \times 0.2 = 0.32$$

Hence

$$P(\text{high quality oil} \mid \text{soil}) = \frac{0.5 \times 0.2}{0.32} = 0.3125$$

Similarly

$$P(\text{medium quality oil} \mid \text{soil}) = \frac{0.8 \times 0.2}{0.32} = 0.50$$

$$P(\text{no oil} \mid \text{soil}) = \frac{0.3 \times 0.2}{0.32} = 0.1875$$

2. *The firm should interpret the soil test as increasing the probability of medium quality oil, but making both no oil and high quality oil less likely than their initial probabilities.*

Problem B. (20 points)
A screening test for high blood pressure, corresponding to a diastolic blood pressure of 90mm Hg or higher, produced the following probability table

	Hypertension	
Test	Present	Absent
+ve	0.09	0.03
-ve	0.03	0.85

1. What's the probability that a random person has hypertension?
2. What's the probability that someone tests positive on the test?
3. Given a person who tests positive, what is the probability that they have hypertension?
4. What would happen to your probability of having hypertension given you tested positive if you initially thought you had a 50% chance of having hypertension.

1. $P(H) = 0.09 + 0.03 = 0.12$
2. $P(+) = 0.09 + 0.03 = 0.12$
3. $P(H|+) = P(+|H)P(H)/P(+) = 0.09/0.12 = 0.75$
4. Using the Bayes rule:
$$P(H|+) = \frac{P(+|H)P(H)}{P(+)}$$
$P(+|H) = 0.09/0.12 = 0.75$, $P(+|\bar{H}) = 0.03/(1 - 0.12) = 0.034$, $P(H) = 0.5$, and $P(\bar{H}) = 0.5$. Finally
$$P(+) = P(+|H)P(H) + P(+|\bar{H})P(\bar{H}) = 0.75 * 0.5 + 0.035 * 0.5 = 0.3925$$
Thus 0.75*0.5/0.3925 = 0.95

Problem B. (20 points)
Suppose that a hypothetical baseball player (call him "Rafael") tests positive for steroids. The test has the following sensitivity and specificity

1. If a player is on Steroids, there's a 95% chance of a positive result.
2. If a player is clean, there's a 10% chance of a positive result.

A respected baseball authority (call him "Bud") claims that 1% of all baseball players use Steriods. Another player (call him "Jose") thinks that there's a 30% chance of all baseball players using Steriods.

- What's Bud's probability that Rafael uses Steroids?
- What's Jose's probability that Rafael uses Steroids?

Explain any probability rules that you use.

Let T and $\bar{T}$ be positive and negative test results. Let S and $\bar{S}$ be using and not using Steriods, respectively. We have the following conditional probabilities

$$P(T|S) = 0.95 \text{ and } P(T|\bar{S}) = 0.10$$

For our prior distributions we have $P_{Bud}(S) = 0.01$ and $P_{Jose}(S) = 0.30$.
From Bayes rule and we have

$$P(S|T) = \frac{P(T|S)P(S)}{P(T)} \tag{1}$$

and by the law of total probability

$$P(T) = P(T|S)P(S) + P(T|\bar{S})P(\bar{S}) \tag{2}$$

Applying these two probability rules gives

$$P_{Bud}(S|T) = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.10 \times 0.99} = 0.0876$$

and

$$P_{Jose}(S|T) = \frac{0.95 \times 0.3}{0.95 \times 0.3 + 0.10 \times 0.7} = 0.8028$$

Problem D. (20 points)

A Breathalyzer test is calibrated so that if it is used on a driver whose blood alcohol concentration exceeds the legal limit, it will read positive 99% of the time, while if the driver is below the limit it will read negative 90% of the time. Suppose that based on prior experience, you have a prior probability that the driver is above the legal limit of 10%.

1. If a driver tests positive, what is the posterior probability that they are above the legal limit?
2. At Christmas 20% of the drivers on the road are above the legal limit. If all drivers were tested, what proportion of those testing positive would actually be above the limit
3. How does your answer to part 1 change. Explain

Let the events be defined as follows:
E - Alcohol concentration exceeds legal limit
NE - Alcohol concentration does not exceed legal limit
P - Breathalyser reads positive
N - Breathalyser reads negative

$$\begin{aligned} P(P|E) &= 0.99 \\ P(N|NE) &= 0.90 \\ P(E) &= 0.10 \end{aligned}$$

Part 1

Based on above, we have $P(P|NE) = 1 - P(N|NE) = 0.10$ and $P(NE) = 1 - P(E) = 0.90$. Then,

$$\begin{aligned} P(E|P) &= \frac{P(P|E)P(E)}{P(P)} \\ &= \frac{P(P|E)P(E)}{P(P|E)P(E) + P(P|NE)P(NE)} \\ &= \frac{0.99 \times 0.10}{0.99 \times 0.10 + 0.10 \times 0.90} \\ &= 52.38\% \end{aligned}$$

Part 2

Now, we have $P(E) = 0.20$. Thus, $P(NE) = 1 - P(E) = 0.80$. We now calculate

$$\begin{aligned} P(E|P) &= \frac{P(P|E)P(E)}{P(P)} \\ &= \frac{P(P|E)P(E)}{P(P|E)P(E) + P(P|NE)P(NE)} \\ &= \frac{0.99 \times 0.20}{0.99 \times 0.20 + 0.10 \times 0.80} \\ &= 71.22\% \end{aligned}$$

Thus, 71.22% of all the drivers who test positive would be above the legal limit.

Part 3

Compared to Part 1, the posterior probability increases due to the fact that the probability of a driver testing positive and exceeding the legal limit increases as does the probability of testing positive. However, the increase in the probability of testing positive and exceeding the legal limit is greater than the increase in the probability of testing positive which results in an increase in the posterior probability.

Problem D. (20 points)

The quality of Nvidia's graphic chips have the probability that a randomly chosen chip being defective is only 0.1%. You have invented a new technology for testing whether a given chip is defective or not. This test will always identify a defective chip as defective and only "falsely" identify a good chip as defective with probability 1%

This give the following probabilities (D represents "defective chip" and T represents "test result indicates defective chip"):

$$P(D) = 0.001$$

$$P(T|D) = 1$$

$$P(T|\bar{D}) = 0.01$$

1. What are the sensitivity and specificity of your testing device?

   Sensitivity: $P(T|D) = 1$

   Specificity: $P(\bar{T}|\bar{D}) = 1 - P(T|\bar{D}) = 1 - 0.01 = 0.99$

2. Given that the test identifies a defective chip, what's the posterior probability that it is actually defective?

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})}$$

$$= \frac{(1)(0.001)}{(1)(0.001) + (0.01)(1 - 0.001)} = 0.090992$$

3. What percentage of the chips will the new technology identify as being defective?

$$P(T) = P(T|D)P(D) + P(T|\bar{D})P(\bar{D})$$

$$= (1)(0.001) + (0.01)(1 - 0.001) = 0.01099$$

4. Should you advise Nvidia to go ahead and implement your testing device? Explain.

   No, only 9% of those chips indicated to be defective by the test will actually be defective. Essentially, we would be throwing away 91% of the chips indicated to be defective even though they are perfectly fine!

Problem C. (20 points)

The Chicago bearcats baseball team plays 60% of its games at night and 40% in the daytime. They win 55% of their night games and only 35% of their day games. You found out the next day that they won their last game

1. What is the probability that the game was played at night
2. What is the marginal probability that they will win their next game?

Explain clearly any rules of probability that you use.

We have the following info:

$$P(Night) = 0.6, \quad P(Day) = 0.4$$

$$P(Win|Night) = 0.55, \quad P(Lose|Night) = 0.45$$
$$P(Win|Day) = 0.35, \quad P(Lose|Day) = 0.65$$

1.

$$P(Night|Win) = \frac{P(Win|Night)P(Night)}{P(Win)}$$
$$= \frac{P(Win|Night)P(Night)}{P(Win|Day)P(Day) + P(Win|Night)P(Night)}$$
$$\frac{0.55 * 0.6}{0.35 * 0.4 + 0.55 * 0.6} = \frac{0.33}{0.47} = 0.7021$$

2.

$$P(Win) = P(Win|Day)P(Day) + P(Win|Night)P(Night)$$
$$= 0.35 * 0.4 + 0.55 * 0.6 = 0.47$$

Problem B. (20 points)
Several spam filters use Bayes rule.
Suppose that you empirically find the following probability table for classifying emails with the phrase "buy now" in their title as either "spam" or "not spam".

	Spam	Not Spam
"buy now"	0.02	0.08
not "buy now"	0.18	0.72

1. What is the probability that you will receive an email with spam?

   Answer: The probability that you will receive an email with spam is:

   $$Pr(Spam) = .02 + .18 = .2$$

2. Suppose that you are given a new email with the phrase "buy now" in its title

   What is the probability that this new email is spam?

   Answer: The posterior probability is

   $$Pr(Spam|buy\ now) = \frac{Pr(Spam\ and\ buy\ now)}{Pr(buy\ now)} = \frac{.02}{.02 + .08} = .2$$

3. Explain clearly any rules of probability that you use.

   Answer: The marginal probability is given by the law of total probability: $P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$. We also use the definition of conditional probability (not Bayes' rule) $Pr(A|B) = \frac{Pr(A\ and\ B)}{Pr(B)}$

Problem B. (20 points)
The Chicago Cubs are having a great season. So far they've won 72 out of the 100 games played so far. You also have the expert opinion of Bob the sports analysis. He tells you that he thinks the Cubs will win. Historically his predictions have a 60% chance of coming true.

- Calculate the probability that the Cubs will win given Bob's prediction
- Suppose you now learn that it's a home game and that the Cubs win 60% of their games at Wrigley field.

  What's you updated probability that the Cubs will win their game?

1. *First we have*

$$P(\text{Win}|\text{Bob}) = \frac{0.72 \times 0.6}{0.72 \times 0.6 + 0.28 \times 0.4} = 0.79$$

2. *Learning the new information, we use the previous posterior as a prior for the next Bayes update*

$$P(\text{Win}|\text{home}, \text{Bob}) = \frac{0.79 \times 0.60}{0.79 \times 0.60 + 0.21 \times 0.40} = 0.85$$

   it's highly likely that the Cubs will win!