Optional Probability Topics

Module 13

Statistics 251: Statistical Methods

Updated 2019

Learning Outcomes

(1) Probabilities when sampling with replacement vs. sampling without replacement
(2) Conditional probability: probability of event A given that B occurred, law of total probability
(3) Bayes' Theorem: prior and posterior probabilities, probability of B given A using Bayes' Theorem
(4) Tree Diagrams: visualization of Bayes' Theorem and conditional probabilities

Probability Rules Reminder

(1) $0 \leq P(A) \leq 1$
(2) $\sum P(A_i) = 1 = S$
(3) $P(A') = 1 - P(A)$
(4) Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(5) Multiplication rule: $P(A \cap B) = P(A)P(B)$ (independent events only)

Sampling with replacement (swr)

When a member of a population is chosen, it is then replaced back into the population and has another chance to be chosen for the sample; probabilities will not change for the second pick, in other words, events are considered to be independent

Think of a deck of 52 playing cards. When one is drawn from the deck, it is observed (recorded the value and suit) and then placed back into the deck of cards (where it could be drawn again at a later time), meaning there are still 52 cards in the deck when we go back to draw another card.

Sampling without replacement (swor)

When a member of a population is chosen, it is not replaced back into the population and no longer has another chance to be chosen for the sample; probabilities will change for the second pick since there are 1 fewer elements to choose from each time one is chosen, the events are considered to be dependent (not indpendent)

Think of a deck of 52 playing cards. When one is drawn from the deck, it is observed (recorded the value and suit) and then it is left out of the deck of cards (it cannot be drawn again at a later time), meaning there are no longer 52 cards in the deck when we go back to draw another card, there are 51 cards left.

swr and swor examples

Think of randomly drawing 2 cards from a standard deck of 52 cards. There are 4 suits, so the probability of any one suit is $\frac{1}{4}$, etc.

swr: Find the probability that you draw 2 hearts (any value, just from the hearts only)

$$P(2hearts) = P(heart\ on\ 1^{st}\ card\ AND\ heart\ on\ 2^{nd}\ card) = P(heart\ on\ 1^{st})P(heart\ on\ 2^{nd})$$

$$= \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{1}{16} = 0.0625$$

The probability of the second card being a heart is the same for one on the first card since we replaced the first card after drawing it and recording it.

swor: Find the probability that you draw 2 hearts (any value, just from the hearts only)

$$P(2hearts) = P(heart\ on\ 1^{st}\ card\ AND\ heart\ on\ 2^{nd}\ card) = P(heart\ on\ 1^{st})P(heart\ on\ 2^{nd})$$

$$= \left(\frac{13}{52}\right)\left(\frac{12}{51}\right) = \frac{156}{2652} \approx 0.0588$$

The probability of the second card being a heart is different for one on the first card since we did not replace the first card after drawing it and recording it, leaving one less heart in the deck (and one less card).

Conditional Probability

(6) Conditional probability
The probability of event A, given that event B has already occurred, is stated as $P(A \text{ given } B)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

This formula can be modified if the conditional probability of one or both complements is required; can also be used to prove independence. If A and B are independent, then $P(A|B) = P(A)$

Probability Rules Update I

(1) $0 \leq P(A) \leq 1$
(2) $\sum P(A_i) = 1 = S$
(3) $P(A') = 1 - P(A)$
(4) Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(5) Multiplication rule: $P(A \cap B) = P(A)P(B)$ (independent events only)
(6) Conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Pay attention to the formulas. They can be modified for finding complements, as well as solving for unknown values with a bit of algebra

Matrix example

Suppose that $P(A) = 0.5$, $P(B) = 0.3$, $P(A \cap B) = 0.2$

Table 3: confusion matrix

	$P(B)$	$P(B')$	
$P(A)$	0.2	0.3	0.5
$P(A')$	0.1	0.4	0.5
	0.3	0.7	1

Matrix calculations

$$P(A|B') = \frac{P(A \cap B')}{P(B')} = \frac{0.3}{0.7} \approx 0.429$$

Are A and B independent? We can use either $P(A \cap B) = P(A)P(B)$ or $P(A|B) = P(A)$ to prove it.

$$P(A \cap B)? = ?P(A)P(B) \Rightarrow 0.2 \neq (0.5)(0.5)$$

$$P(A|B)? = ?P(A) \Rightarrow 0.429 \neq 0.5$$

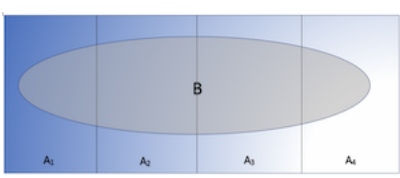
With either method, the results are the same. Since the statements were false, events A and B are not independent (they are dependent).

Law of Total Probability I

Reviewing Rule (6), conditional probability: the probability of event A, given that event B has already occurred:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \; P(B) > 0$$

Many times the probability of interest is the denominator $P(B)$, but it is not directly known. When B arises in connection with events $A_1, A_2, \ldots, A_k$, which constitute a *partition* of the sample space (i.e., they are mutually exclusive (disjoint) and make up the entire sample space); see figure.



Law of Total Probability II

If the probability of each A_i and the conditional probability of B given each A_i are all known, the Law of Total Probability expresses the probability of B as, and holds for any event B:

$$P(B) = \sum_{j=1}^{k} P(B|A_j) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)$$

The events of $A_1, A_2, \ldots, A_k$ can also be thought of as a stratification of the population.

Bayes' Theorem is to probability as Pythagorem's Theorem is to geometry[1]

Bayes' Theorem

(7) Bayes' theorem: Let $A_1, A_2, \ldots, A_k$ be a collection of k mutually exclusive (disjoint) and exhaustive events with *prior* probabilities $P(A_i)$ $(i = 1, 2, \ldots, k)$. Then for any other event B for which $P(B) > 0$, the *posterior* probability of A_j given that B has occurred is

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum P(B|A_j)P(A_j)}$$

Another way to think about it: There is a partition $A_1, A_2, \ldots, A_k$ of the sample space and an event B (as shown in previous figure). The probabilities of the events A_i are given, and so are the conditional probabilities of B given that an A_i has occurred. Bayes' theorem answers the question: Given that B has occurred, what is the probability that a particular A_j has occurred?

(For just 2 events A and B: $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A')P(A')}$)

Bayes' theorem example I

Suppose that a test for using a particular drug has a 99% sensitivity rate and a 99% specificity rate. That is, the test will produce 99% true positive results for drug users and 99% true negative results for non-drug users. Suppose that 0.5% of people are users of the drug. What is the probability that a randomly selected individual with a positive test is actually a drug user?

$P(user) = 0.005$, $P(user') = 1 - P(user) = 1 - 0.005$, $P(+|user) = 0.99$, $P(-|user') = 0.99$, $P(+|user') = 1 - P(-|user) = 1 - 0.99 = 0.01$, and $P(-|user) = 1 - P(-|user') = 1 - 0.99 = 0.01$

$P(user|+) = \frac{P(user\ \cap +)}{P(+)} = \frac{P(+|user)P(user)}{P(+|user)P(user)+P(+|user')P(user')} = \frac{(0.99)(0.005)}{(0.99)(0.005)+(0.01)(0.995)} = 0.332215$

Interpretation of drug use and tests

Even if an individual tests positive, it is more likely that they do not use the drug than that they do. This is because the number of non-users is large compared to the number of users. The number of false positives outweighs the number of true positives. For example, if 1000 individuals are tested, there are expected to be 995 non-users and 5 users. From the 995 non-users, $0.01(995) \approx 10$ false positives are expected. From the 5 users, $(0.99)(5) \approx 5$ true positives are expected. Out of 15 positive results, only 5 are genuine. Yikes!

[1] Jeffreys, Harold (1973). Scientific Inference (3rd ed.). Cambridge University Press. p. 31. ISBN 978-0-521-18078-8.

Bayes' theorem example II

The entire output of a factory is produced on three machines. The three machines account for 20%, 30%, and 50% of the factory output. The fraction of defective items produced is 5% for the first machine, 3% for the second machine, and 1% for the third machine. If an item is chosen at random from the total output and is found to be defective, what is the probability that it was produced by the third machine?

Here, the answer can be reached without recourse to the formula by applying the conditions to any hypothetical number of cases. For example, if 100,000 items are produced by the factory, 20,000 will be produced by Machine A, 30,000 by Machine B, and 50,000 by Machine C. Machine A will produce 1000 defective items, Machine B 900, and Machine C 500. Of the total 2400 defective items, only 500, or 5/24 were produced by Machine C.

Bayes' example II solution

A solution is as follows. Let X_i denote the event that a randomly chosen item was made by the i^{th} machine (for $i = A, B, C$). Let Y denote the event that a randomly chosen item is defective. Then, we are given the following information:

$P(X_A) = 0.2$, $P(X_B) = 0.3$, $P(X_C) = 0.5$, $P(Y|X_A) = 0.05$, $P(Y|X_B) = 0.03$, $P(Y|X_C) = 0.01$

$P(X_C|Y) = \frac{P(X_C \cap Y)}{P(Y)}$. First find $P(Y)$ using the Law of Total Probability.

$P(Y) = P(Y|X_A)P(X_A)+P(Y|X_B)P(X_B)+P(Y|X_C)P(X_C) = (0.05)(0.2)+(0.03)(0.3)+(0.01)(0.5) = 0.024$

Now finish the problem: $P(X_C|Y) = \frac{P(X_C \cap Y)}{P(Y)} = \frac{P(Y|X_C)P(X_C)}{P(Y)} = \frac{(0.01)(0.5)}{0.024} = 0.2083 = \frac{5}{24}$

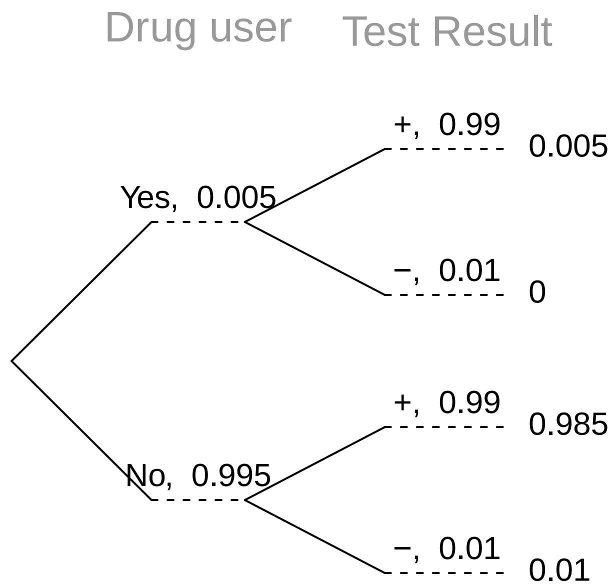
Interpretation of machines and defective products

Given that the item is defective, the probability that it was made by the third machine is only 5/24. Although machine C produces half of the total output, it produces a much smaller fraction of the defective items. Hence the knowledge that the item selected was defective enables us to replace the prior probability $P(X_C) = 0.5$ by the smaller posterior probability $P(X_C|Y) = 0.2083$.

Tree diagrams

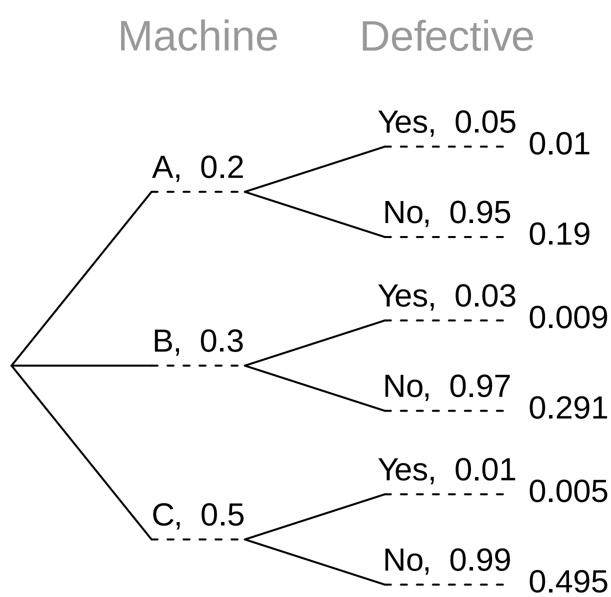
Another way to describe and visualize Bayes' theorem is to use a Tree diagram. Start with the initial condition branches, then the next branches are the conditional probabilities of the next even given the initial conditions. From there, most often the outside edge of the tree has the intersections computed by the product of the initial conditions and conditional probabilities.

Tree for drug use and tests



$P(No\ and\ +) = (0.995)(0.99) = 0.985$, etc.

Tree for defects and machines



Final Probability Rules Update

(1) $0 \leq P(A) \leq 1$
(2) $\sum P(A_i) = 1 = S$
(3) $P(A') = 1 - P(A)$
(4) Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(5) Multiplication rule: $P(A \cap B) = P(A)P(B)$ (independent events only)
(6) Conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$
(7) Bayes' Theorem: $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum P(B|A_j)P(A_j)}$