Date: Jan 18, 2022

Predictive value positive (PV+)

The predictive value positive (PV+) of a screening test is the probability that a person truly has a disease given that the test is positive. That is,

$$PV+= Pr(D^+|T^+)$$

What does the PV+ imply?

How worried a subject with the positive test be?

Predictive value negative (PV-)

The predictive value negative (PV-) of a screening test is the probability that a person truly does not have a disease given that the test is negative. That is,

$$PV-= Pr(D^-|T^-)$$

What does the PV- imply?

How reassured a subject with the negative test be?

Example 3.23 text book pg. 55

Suppose that among 100,000 women with negative mammograms (T^-), 20 will be diagnosed with breast cancer (D^+) within 2 years, i.e., $Pr(D^+|T^-) = \frac{20}{100000} = 0.0002$, whereas 1 woman in 10 with positive mammograms will be diagnosed with breast cancer within 2 years, i.e., $Pr(D^+|T^+) = \frac{1}{10} = 0.1$.

(a) Find PV^+ and interpret the result.

(b) Find PV^- and interpret the result.

Solution

(a) $PV^+ = Pr(D^+|T^+) = 0.1$.

This result suggests that if the mammogram is positive, the woman has a 10% chance of developing breast cancer (PV^+ = .10).

(b) $PV^- = Pr(D^-|T^-) = 1 - Pr(D^+|T^-) = 1 - 0.0002 = 0.9998$.

This result suggests that if the mammogram is negative, the woman is virtually certain *not* to develop breast cancer over the next 2 years ($PV^- \approx 1$).

False positive and false negative

A **false positive** of a test is the probability of having a positive test result given no disease. That is,

$$False\ positive = Pr(T^+|D^-)$$

A **false negative** of a test is the probability of having a negative test result given disease. That is,

$$False\ positive = Pr(T^-|D^+)$$

A test is effective if it has **low** false positive and false negative values.

Measure of Effectiveness of a Test

Sensitivity

The sensitivity of a test is the probability of having a positive test result given disease. That is

$$Sensitivity = Pr(T^+|D^+)$$

Specificity

The specificity of a test is the probability of having a negative test result given no disease. That is

$$Specificity = Pr(T^-|D^-)$$

A test is effective if it has **high** sensitivity and specificity values.

Example 3.11 pg. 57

The level of ***prostate-specific antigen*** (PSA) in the blood is frequently used as a screening test for prostate cancer. The value of PSA $\geq$ 4.1 ng/dL is considered a positive test (T+) (ng/dL refers to **nanograms per deciliter**). The following table summarizes the relationship between a positive PSA test (≥4.1 ng/dL) and prostate cancer.

Table of T by D			
T	**D**		
	+	**-**	**Total**
+	92	27	119
-	46	72	118
Total	138	99	237

Compute the following:

(a) False positive of PSA test

(b) False negative of PSA test

(c) Sensitivity of PSA test

(d) Specificity of PSA test

Solution

(a) False positive of PSA test $= Pr(T^+|D^-) = \frac{27}{99} = 0.273$

(b) False positive of PSA test $= Pr(T^-|D^+) = \frac{46}{138} = 0.333$

(c) Sensitivity of PSA test$= Pr(T^+|D^+) = \frac{92}{138} = 0.667$

(d) Specificity of PSA test$= Pr(T^-|D^-) = \frac{72}{99} = 0.727$

Given the assumption that the table summarizing the relationship between PSA test and prostate cancer represents the true presence of prostate cancer.

(e) Evaluate *PV+* $= Pr(D^+|T^+) = \frac{92}{119} = 0.773$

(f) Evaluate *PV-* $= Pr(D^-|T^-) = \frac{72}{118} = 0.610$

However, if the assumption made in (e)-(f) is not true, we need to use Bayes' rule to compute PV+/-.

$$P(A|B) = P(A \cap B)/P(B)$$

Multiplicative rule for any two events

For any two events A and B, using the definition of conditional probability, it follows that

$$P(A \cap B) = P(A|B)P(B)$$

or equivalently,

$$P(A \cap B) = P(B|A)P(A)$$

The above results are what we call multiplicative rules for two events A and B.

Thus,

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Law of total probability

$$\bar{A} = A^c, to\ refer\ to\ complement\ of\ the\ event\ A$$

Equation 3.6: For two events A and B

$$P(B) = P(B|A) \times P(A) + P(B\ |\bar{A}) \times P(\bar{A})$$

This formula enables us to compute the unconditional probability of B using the sum of the conditional probability of B given other events, weighted by the probability of the other events.

Proof: $B = (B \cap A) \cup (B \cap \bar{A})$

Also,$(B \cap A) \cap (B \cap \bar{A}) = \emptyset$

Thus,

$P(B) = P(B \cap A) + P(B \cap \bar{A})$, by sum of mutually exclusive events

$= P(B|A) \times P(A) + P(B\ |\bar{A}) \times P(\bar{A})$, by multiplicative law of probability

Example 3.23 text book pg. 55

Suppose that among 100,000 women with negative mammograms, 20 will be diagnosed with breast cancer within 2 years, whereas 1 woman in 10 with positive mammograms will be diagnosed with breast cancer within 2 years. Suppose that 7% of the general population of women have a positive diagnosis.

Given the scenario, find the probability of developing breast cancer among the women in the general population over the next 2 years.

Solution:

Let us define the events:

B: Breast can positive in the next two year.

M^+: Mammogram test is positive

M^-: Mammogram test is negative

We have, $P(B|M^-) = \frac{20}{100000} = 0.0002$, $P(B|M^+) = \frac{1}{10} = 0.1$ and $P(M^+) = 0.07$.

So, $P(M^-) = 1 - P(M^+) = 0.93$. Then, by the LTP,

$$P(B) = P(B|M^-) \times P(M^-) + P(B|M^+) \times P(M^+) = 0.0002 * 0.93 + 0.1 * 0.07 = 0.00719$$

Extension of Law of total probability for more than two events

Equation 3.7: If $A_1, A_2, \dots, A_k$ are k mutually exclusive and exhaustive events in the sample space, then

$$P(A) = \sum_{i=1}^{k} P(A \cap A_i) = \sum_{i=1}^{k} P(A|A_i) \times P(A_i)$$

This formula enables us to compute the unconditional probability of A using the sum of the conditional probability of A given A_i, weighted by $P(A_i)$.

Proof: $A = \cup_{i=1}^{k}(A \cap A_i) = (A \cap A_1) \cup (A \cap A_2) \cup \dots (A \cap A_k)$

Also, $(A \cap A_i) \cap (A \cap A_j) = \emptyset$ for all $i \neq j = 1, 2, \dots, k$

Thus, $P(A) = \sum_{i=1}^{k} P(A \cap A_i)$, by sum of mutually exclusive events

$= \sum_{i=1}^{k} P(A|A_i) \times P(A_i)$, by the definition of the multiplicative law of probability.

Example 3.22 Pg. 54 Text book

A 5-year study is conducted for a population of 5000 people 60 years of age or older to evaluate how many people will develop a cataract over the next 5 years. From a census data it appears that 45% of this population is 60-64 years of age, 28% are 65-69 years of age, 20% are 70-74 years of age, and 7% are 75 or older. An existing eye study reveals that 2.4%, 4.6%, 8.8% and 15.3% of the people in these respective age groups will develop a cataract over the next five years.

(a) What is the probability that a randomly selected person from this population will develop a cataract?

(b) What is an estimate of the number of people who will have a cataract?

Solution:

Define the following events:

$A_i = \{people\ in\ ith\ age\ group\}, i = 1,2,3,4.$
$B = \{develop\ cataract\ in\ the\ next\ five\ year\}$

Then, we have,
$P(A_i) = 0.45, 0.28, 0.20, 0.07$ for $i =$ 1,2,3,4, respectively.
$P(B|A_i) = 0.024, 0.046, 0.088, 0.153$ for $i =$ 1,2,3,4, respectively

(a) $P(B) = \sum_{i=1}^{4} P(B|A_i) \times P(A_i)$

$= 0.024 * 0.45 + 0.046 * 0.28 + 0.088 * 0.20 + 0.153 * 0.07 = 0.052$

(b) The number of people with cataract in the next five year$= 5000 \times 0.052 = 260$

Bayes' Theorem

Given k mutually exclusive and exhaustive events $A_1, A_2, \dots, A_k$ of the sample space S. Let A be any event in the sample space. Then,

$$P(A_i|A) = \frac{P(A|A_i) \times P(A_i)}{\sum_{i=1}^{k} P(A|A_i) \times P(A_i)}$$

Proof: By the definition of the conditional probability

$$P(A_i|A) = \frac{P(A \cap A_i)}{P(A)} = \frac{P(A|A_i) \times P(A_i)}{P(A)} = \frac{P(A|A_i) \times P(A_i)}{\sum_{i=1}^{k} P(A|A_i) \times P(A_i)}$$

Example 3.22 follow-up for Bay's Theorem Pg. 54 Text book

From a census of a population aged 60 years or above, it appears that 45% of the population is 60-64 years of age, 28% are 65-69 years of age, 20% are 70-74 years of age, and 7% are 75 or older. It also appears that 2.4%, 4.6%, 8.8% and 15.3% of the people in these respective age groups will develop a cataract over the next five years.

Define the following events:

$A_i = \{people\ in\ ith\ age\ group\}, i = 1,2,3,4.$

$B = \{develop\ cataract\ in\ the\ next\ five\ year\}$

(c) Express all the percentage expressions such as 45%, 28%,…,15.3% as probability of related events.

(d) What is the probability that a randomly selected person from this population will develop a cataract?

(e) Given that a randomly selected person will a cataract in the next five year, what is the probability that this person belongs to age-group

(i) 60-64 years?

(ii) 70-44 years?

Solution:

(a) $P(A_i) = 0.45, 0.28, 0.20, 0.07$ for $i = 1, 2, 3, 4,$ respectively, and $P(B|A_i) = 0.024, 0.046, 0.088, 0.153$ for $i = 1,2,3,4$, respectively

(b) $P(B) = \sum_{i=1}^{4} P(B|A_i) \times P(A_i)$

$$= 0.024 * 0.45 + 0.046 * 0.28 + 0.088 * 0.20 + 0.153 * 0.07 = 0.052$$

(c) $P(A_i|B) = \frac{P(B|A_i) \times P(A_i)}{\sum_{i=1}^{4} P(B|A_i) \times P(A_i)}$. So,

(i) $P(A_1|B) = \frac{P(B|A_i) \times P(A_1)}{\sum_{i=1}^{4} P(B|A_i) \times P(A_i)} = \frac{0.024 * 0.45}{0.052} = 0.208$

(ii) $P(A_3|B) = \frac{P(B|A_3) \times P(A_3)}{\sum_{i=1}^{4} P(B|A_i) \times P(A_i)} = \frac{0.088 * 0.20}{0.052} = 0.3385$

Definition 3.18 pg. 62: ROC curve

A **receiver operating characteristic (ROC) curve** is a plot of the sensitivity (on the y-axis) versus (1 – specificity) (on the x-axis) of a screening test, where the different points on the curve correspond to different cutoff points used to designate test-positive. The best cut-off has the highest true positive rate together with the lowest false positive rate.

The area under the ROC curve gives an idea about the benefit of using the test(s) in question.

Suggested HW 8th ed 3.1-3.27 text