Probability

Readings: Ch 14, 15 DVB

What is probability?

- A number between 0 and 1 assigned to *events*
- Event: an event is a set of outcomes of an experiment to which a probability is assigned. It is the result of an observation or experiment, or the description of some potential outcome.
- Long run relative frequency interpretation.

$$P(A) = \frac{number\ of\ outcomes\ in\ A}{number\ of\ possible\ outcomes}$$

  - Example: $A =$ female baby

$$P(A) = \frac{no.\ female\ babies}{no.\ babies}$$

Operations on events and probability

- *Complement*, A^c or $\bar{A}$, the event that A does not occur.
  - Example: $A^c = male\ baby$
- *Union*, $A \cup B$, either A or B or both occur.
  - Example: $B = healthy\ baby$,
    $A \cup B = baby\ is\ female\ or\ healthy\ or\ both$
- *Intersection*, $A \cap B$, both A and B occur.
  - Example: $A \cap B = baby\ is\ female\ and\ healthy$
- *Sample space*, S, union of all events, must occur, $P(S) = 1$.
- *Null event*, ϕ, can't occur, $P(\phi) = 0$.
- A and B are *mutually exclusive* or *disjoint* if they can't both occur, ie if $A \cap B = \phi$.
  - Example: $A = pill\ is\ red$, $B = pill\ is\ green$
- A and B are *independent* if occurrence of one doesn't affect the probability of the other.
  - Example: $A = first\ baby\ is\ female$, $B = second\ baby\ is\ male$

Additive Rule

- If A and B are mutually exclusive (or disjoint),

$$P(A \cup B) = P(A) + P(B).$$

  - Example: A box contains 3 pills, one is red (R), one is green (G) and one is white (W). If you choose one pill *at random*, then the probability it is either red or green is

$$P(R \cup G) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

- Otherwise

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

  - Example: If you choose two pills *at random* from the same box, then there are 3 possible outcomes, all equally likely, S = RG, RW, GW, so $P(RG) = \frac{1}{3}$, $P(R) = \frac{2}{3} = P(G)$, and

$$P(R \cup G) = \frac{2}{3} + \frac{2}{3} - \frac{1}{3} = \frac{3}{3} = 1.$$

- Because A and $\bar{A}$ are mutually exclusive and $A \cup \bar{A} = S$, $P(\bar{A}) = 1 - P(A)$.
  - Sometimes it is much easier to calculate $P(\bar{A})$ than $P(A)$.

Conditional Probability

- The *conditional probability of B given A* is

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

  - Example: Choosing two from the box with one red one green and one white pill

$$P(R|G) = \frac{P(R \cap G)}{P(G)} = \frac{1}{3}/\frac{2}{3} = \frac{1}{2}.$$

Independence

- A and B are *independent* if and only if $P(B|A) = P(B)$ or, equivalently, if $P(A|B) = P(A)$.
  - Example: When choosing two from the box, the events R and G are not independent, because $P(R|G) = \frac{1}{2}$ is not the same as $P(R) = \frac{2}{3}$.

Multiplicative Rule

- In general,
$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B).$$
- If A and B are independent, then
$$P(A \cap B) = P(A)P(B)$$

Example: Consider the following table of probabilities:

		Cured		
		Yes	No	Total
Side Effects	Yes	.1	.1	.2
	No	.6	.2	.8
	Total	.7	.3	1.0

- Let A be the event that a patient has side effects, and B be the event that they are cured.
- Then $A \cup B$ is the event that a patient has side effects or is cured or both.
- $A \cap B$ is the event that a patient has side effects and is cured.
- A and B are not mutually exclusive because it is possible they have side effects and be cured ($A \cap B \neq \phi$).
- From the table we can conclude
  1. $P(A) = .2$, $P(B) = .7$
  2. $P(A \cap B) = .1$, A and B are not mutually exclusive.
  3. $P(A \cup B) = .1 + .1 + .6 = .8$
  4. $P(no\ side\ effects\ or\ cured) =$
  5. $P(A \cap B) = .1$, but $P(A)P(B) = .2 \times .7 = .14$ so A and B are not independent.
  6. $P(A|B) = P(A \cap B)/P(B) = .1/.7 = .14$ (which is $< P(A)$)
  7. $P(B|A) = P(A \cap B)/P(A) = .1/.2 = .5$ (which is $< P(B)$)

Diagnostic Testing

- D^+ is event subject has disease, $P(D^+)$ is **prevalence**
- D^- is event subject does not have disease, $P(D^-) = 1 - P(D^+)$
- T^+ is event subject has positive test result
- $P(T^+|D^+)$ is the *test* **sensitivity**
- T^- is event subject has negative test result
- $P(T^-|D^-)$ is the *test* **specificity**

Q: What is the probability of a positive test result?

$$P(T^+) = P(T^+|D^+)P(D^+) + P(T^+|D^-)P(D^-)$$

- This is an application of the *law of total probability*.

Q: What is the probability someone with a positive test result has the disease?

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

- This is the *predictive value* of a positive test or *positive predictive value* of the test.
- This is an application of *Bayes Rule*.

Q. What is the probability someone with a negative test result does not have the disease?

$$P(D^-|T^-) = \frac{P(D^- \cap T^-)}{P(T^-)} = \frac{P(T^-|D^-)P(D^-)}{P(T^-)}$$

- This is the *negative predictive value* of the test.

Bayes' theorem: For two events A and B,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^-)P(A^-)}$$

Example: The Pap smear is used as a test for cervical cancer. Studies have shown that the probability of a false negative is .1625, and that the probability of a false positive is .1864. The prevalence of cervical cancer is .000083.

From the statement $P(D+) = .000083$, $P(T + |D-) = .1864$, $P(T - |D+) = .1625$.

1. What is the sensitivity?

   P(false negative) $= P(T - |D+) =$ .1625
   sensitivity$=P(T + |D+) = 1 - P(T - |D+) = 1 - .1625.$

2. What is the specificity?

   P(false positive) $= P(T + |D-) = .1864$

   specificity $= P(T - |D-) = 1 - P(T + |D-) = 1 - .1864$

3. What is the probability of testing positive?
   $P(T+) = P(T + |D+)P(D+) + P(T + |D-)P(D-) =$

   $(1 - .1625) * .000083 + .1864 * (1 - .000083)$

4. What is the **positive predictive value**? using Bayes' theorem,

   $ppv = P(D + |T+) = P(T + |D+)P(D+)/P(T+)$

   all pieces on right hand side have already been calculated, or are given.

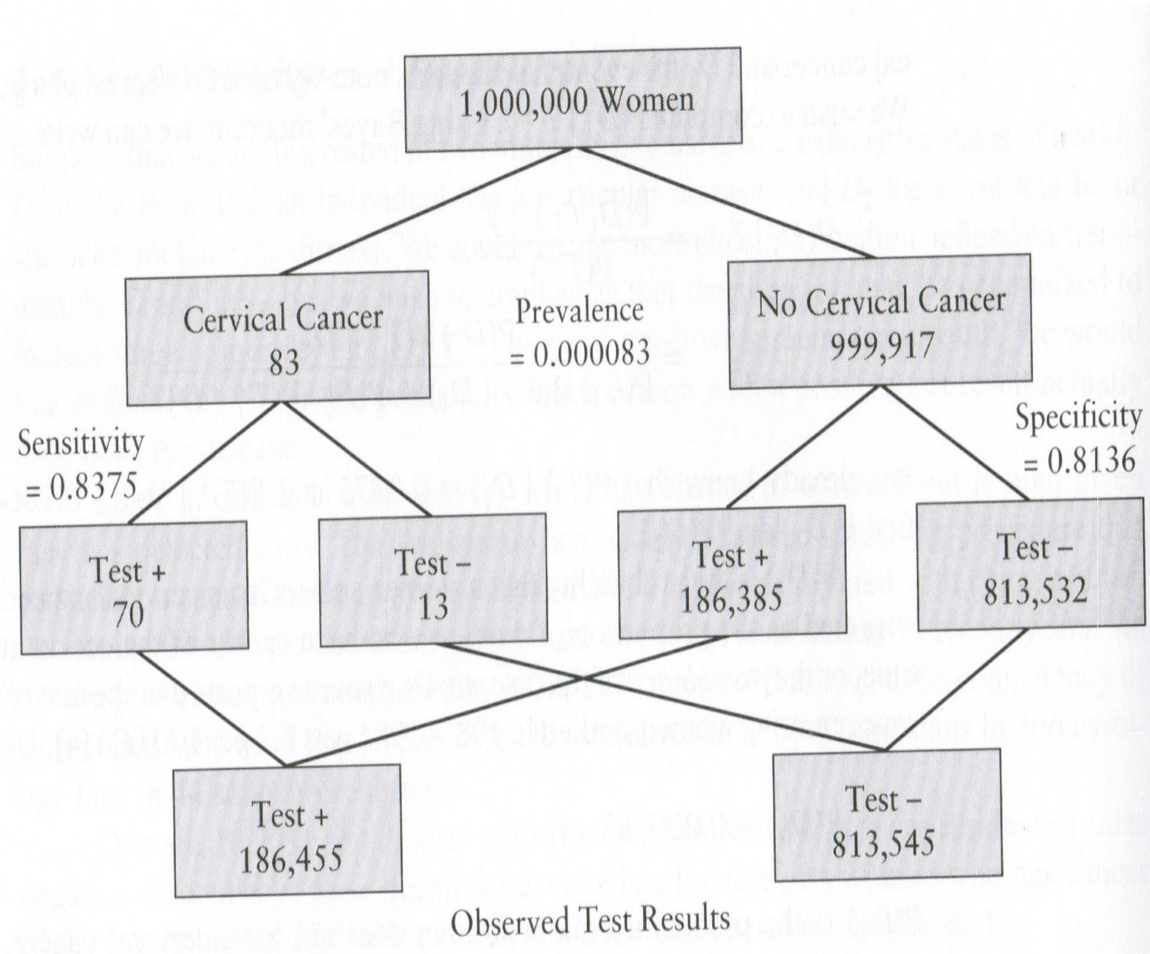
5. What is the negative predictive value?

   using Bayes' theorem, $npv = P(D - |T-) = P(T - |D-)P(D-)/P(T-)$
   all three terms on right hand side have been calculated, are given, or are easy to find. for example $P(T-) = 1 - P(T+)$.

(answers on next page show some additional calculations)

- $P(T^-|D^+) = .1625$ (probability of false negative [test negative given diseased] GIVEN)
- $P(T^+|D^+) = 1 - .1625 = 0.8375$ (sensitivity)
- $P(T^+|D^-) = .1864$ (probability of false positive [test positive given non-diseased] GIVEN)
- $P(T^-|D^-) = 1 - P(T^+|D^-) = 0.8136$ (specificity)
- $P(D^+) = .000083$ (prevalence [probability diseased] GIVEN)
- $P(D^-) = 1 - P(D^+) = 0.999917$ (probability non-diseased)
- $P(T^+ \cap D^+) = P(T^+|D^+)P(D^+) = (1 - .1625) * .000083 = 6.95125e - 05$ (prob. test positive and diseased)
- $P(T^+ \cap D^-) = .1864 * (1 - .000083) = 0.1863845$ (prob. test positive and non-diseased)
- $P(T^- \cap D^+) = .1625 * .000083 = 1.34875e - 05$ (prob. test negative and diseased)
- $P(T^- \cap D^-) = (1 - .1864) * .999917 = 0.8135325$ (prob. test negative and non-diseased)
- $P(T^+) = P(T^+ \cap D^-) + P(T^+ \cap D^+) = 0.1864540$ (probability of positive test)
- $P(T^-) = 1 - P(T^+) = 0.813546$ (probability of negative test)
- $ppv = P(T^+ \cap D^+)/P(T^+) = 0.0003728131$ (positive predictive value = prob. diseased given test positive)
- $npv = P(T^- \cap D^-)/P(T^-) = 0.9999834$ (negative predictive value = prob. non-diseased given test negative)

The following figure uses a **tree diagram** to show the performance of the Pap smear as a diagnostic test for cervical cancer.



(from Pagano & Gauvreau: Principles of Biostatistics, 2nd edition, page 138)