Topic 12 - Bayes' Theorem

Statistics for Managers

June 3, 1999

Thomas Bayes (1702-61) discovered an approach to statistical inference that was far more advanced than the traditional thinking of the mathematicians of the time. Up to the time of Bayes, the focus of mathematicians was on the behavior of samples from known population, but Bayes reversed the idea to determine the properties of a population based on a sample. In "An Essay Towards the Solving a Problem in the Doctrines of Chance," he presented as "Proposition 9," what is now known as *Bayes' theorem*. The essay is, perhaps, one of the least understood but most famous and controversial contributions in the history of science. In modern times, Bayes was rediscovered and his theorem has laid the foundations for modern decision making.

Bayes Theorem

Let A and B be two dependent events, then

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

Proof:

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

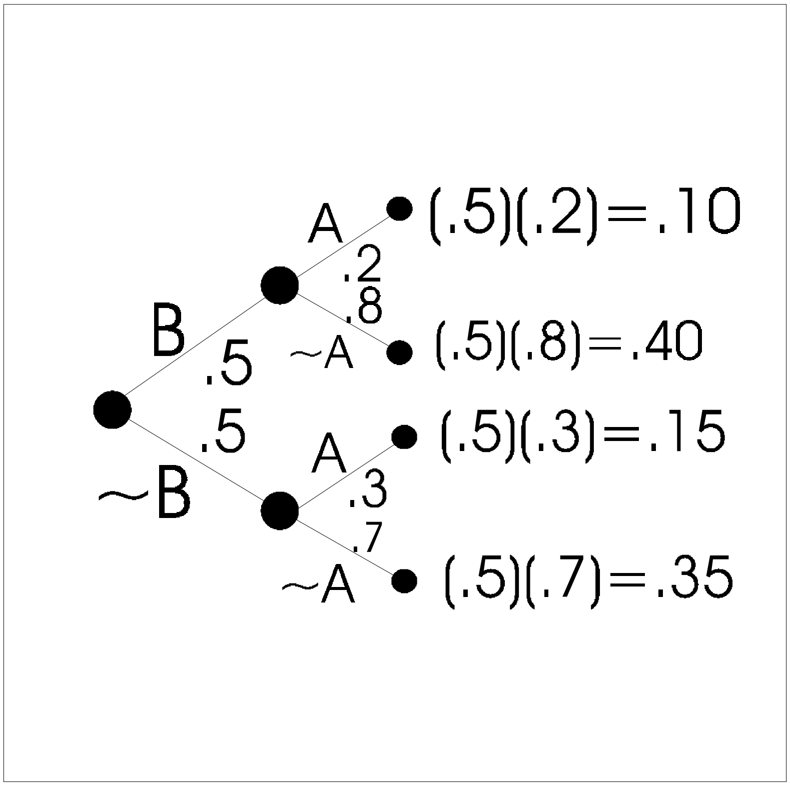
But

$$P(A) = P(A \cap B) + P(A \cap \overline{B})$$

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

Upon substituting the expansion of P(A), we have the final formula.

Example 1. Using odds and Bayes Theorem



$oddsB|A = 10{:}15$

$$P(B|A) = \frac{10}{10+15} = .40$$

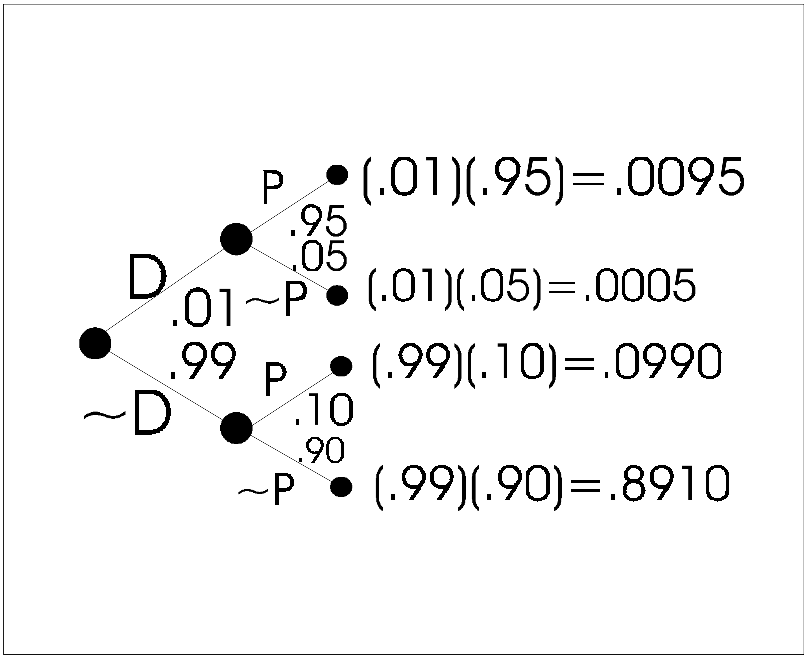
$oddsB|\overline{A} = 40{:}35$

$$P(B|\overline{A}) = \frac{40}{40+35} = .53$$

In the above example, suppose you find out that A occurred, what would be your best guess: B or ~B occurred? How about if ~A occurred?

Example 2. Disease screening

This example shows to exercise caution with disease screening in a general population with a small proportion of subjects having the disease: D = Disease, P = Positive Test Results. For some diseases and medical tests, these probabilities are reasonable: Pr (P|D) = 0.95, Pr(P|~D) = 0.10, Pr(D) = 0.01. Using the probabilities, we construct the following tree for application of Bayes Theorem:



$$oddsD|P = 95{:}990$$

$$P(D|P) = \frac{95}{95+990} = .0876$$

$$oddsD|\overline{P} = 5{:}8910$$

$$P(D|\overline{P}) = \frac{5}{5+8910} = .0006$$

Therefore, no matter what a testing result is, positive (P) or negative (~P), the probability of actually having the disease is small. How accurate should the drug test be for it to be useful in a general population? Another misapplication of general disease screening is drug testing in schools and in work places.

Example 3. Bayesian Statistical Hypothesis Testing.

A functioning manufacturing process should produce computer hardware components with only 0.1 percent defective. A certain components reseller is getting more defective returns than 0.1 percent would indicate, and they suspect the rate is more like one percent. Their experience after selling 1000 parts is five defective returns. We can test which of two Binomial populations, BIN (1000, 0.001) or BIN (1000, 0.01) is more likely; the former being the manufacturer claim of reliability and the latter being the reseller's suspicion.

In statistical terms, we have two hypotheses to explain the true population:

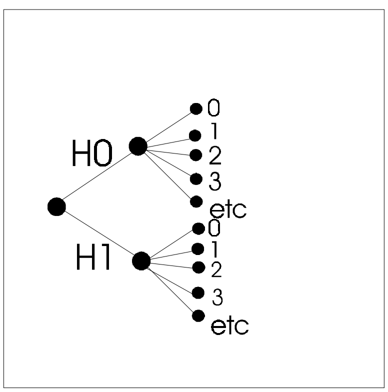
- A *null* hypothesis represents the status quo, or in other words, a state of the population requiring no remedial action.
- The *alternate* hypothesis represents a state of the population that must be corrected or otherwise attended to; in other words, the investigators take some kind of action.
- The null hypothesis is represented by H_0 and the alternative hypothesis is represented by H_1.

For our problem we have:

$$H_0: BIN(1000,.001)$$
$$H_1: BIN(1000,.01)$$

Using the Poisson approximation to the Binomial, these hypotheses may be replaced by:

$$H_0: POISSON(1)$$
$$H_1: POISSON(10)$$



We need to calculate the probability of the branches: $P(H_0)P(Y = 5|H_0)$ and $P(H_1)P(Y = 5|H_1)$:

$$P(H_0)P(Y=5 \mid H_0) = P(H_0)\frac{(1)^5 e^{-1}}{5!} = .0031P(H_0)$$
$$P(H_1)P(Y=5 \mid H_1) = P(H_1)\frac{(10)^5 e^{-10}}{5!} = .0378P(H_1)$$

Now assume $P(H_0) = 0.5$ and $P(H_1) = 0.5$, since we are not in favor of either hypothesis. We calculate the probability of each hypothesis given Y=5:

$$odds(H_0 \mid Y = 5) = 31:378$$

$$P(H_0 \mid Y = 5) = \frac{31}{31 + 378} = .075$$

$$odds(H_1 \mid Y = 5) = 378:31$$

$$P(H_1 \mid Y = 5) = \frac{378}{31 + 378} = .925$$

Conclusion: Accept H_1 since it has the largest probability. Thus, the investigator should act to inform the manufacturer there is a problem with the manufacturing process, and the reseller will not place anymore orders until the problem is remedied.

• Table 1 Table of Poisson Probabilities for H0: μ = 1 and H1: μ = 10

y	P(Y = y|H0)	P(Y = y|H1)	P(H0|Y = y)	P(H1|Y = y)
0	0.3679	0.0000	0.9999	0.0001
1	0.3679	0.0005	0.9988	0.0012
2	0.1839	0.0023	0.9878	0.0122
3	0.0613	0.0076	0.8901	0.1099
4	0.0153	0.0189	0.4476	0.5524
5	0.0031	0.0378	0.0750	0.9250
6	0.0005	0.0631	0.0080	0.9920
7	0.0001	0.0901	0.0008	0.9992
8	0.0000	0.1126	0.0001	0.9999
9	0.0000	0.1251	0.0000	1.0000
10	0.0000	0.1251	0.0000	1.0000
11	0.0000	0.1137	0.0000	1.0000
12	0.0000	0.0948	0.0000	1.0000
13	0.0000	0.0729	0.0000	1.0000
14	0.0000	0.0521	0.0000	1.0000
15	0.0000	0.0347	0.0000	1.0000
16	0.0000	0.0217	0.0000	1.0000
17	0.0000	0.0128	0.0000	1.0000
18	0.0000	0.0071	0.0000	1.0000