Bayes Theorem

Example: A skin test to TB is 97% accurate in the sense that someone with TB has a 97% of testing positive (3% chance of a false negative). It is 98% accurate in the sense that someone without TB has a 98% chance of testing negative (2% chance of a false positive). In the population of interest, 1 in 20,000 people has TB.

Question: If you test positive what are the chances that you have TB?

Events:
POS – testing positive
NEG – testing negative
D – has disease

\[
P(D|POS) = \frac{P(D \cap POS)}{P(POS)} = \frac{(P(POS|D)P(D))}{P(POS)}
\]

\[
P(POS|D) = 0.97 \quad P(NEG|D) = 0.03
\]

\[
P(POS|D)^c = 0.02 \quad P(NEG|D)^c = 0.98
\]

\[
P(D) = 0.00005 \quad P(D)^c = 0.99995
\]

\[
P(D|POS) = \frac{(0.97*0.00005)}{(0.97*0.00005+0.02*0.99995)} = 0.0024
\]

Independence

Definition: Two events A and B are called independent IFF their joint probability is equal to the product of their individual probabilities:

\[
P(A \cap B) = P(A)*P(B)
\]

Otherwise, we say that A and B are dependent.

There is no way to determine whether two events are independent or dependent other than to compute the three numbers and compare them.

Note: If A and B are independent, then assuming \( P(A) \neq 0 \neq P(B) \)

\[
P(A|B) = P(A)\]

\[
P(B|A) = P(B)
\]

Example: Toss a fair coin 10 times

\[E = \text{Heads on the first 9 tosses}\]
\[F = \text{Heads on the last toss}\]

\[
P(E \cap F) = P(E)P(F) ?
\]

\[1/1024 = (2/1024)(512/1024)\]
What if you have 3 or more events?

**Definition:** Suppose $A_1, A_2, A_3, \ldots, A_n$ are events, then we say they are independent IFF

\[ P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n) = \prod_{i=1}^{n} P(A_i) \]

This implies pairwise independence, but not conversely.

**Note:** This is stronger than saying they are pairwise independent. $A_i$ and $A_j$ are independent for $i \neq j$.

**Second Order Linear Difference Equations with Constant Coefficients**

\[ p_{k+1} + a p_k + b p_{k-1} = c_k \]

$a$, $b$ are scalars (constant coefficients) and $c$ can vary with $k$

If $c_k = 0$, $p_{k+1} + a p_k + b p_{k-1} = 0$ is a homogenous equation

Initial conditions $p_0$ and $p_1$ are specified.

A typical solution has the form:

\[ u_k = m^k \text{ for some } m \]

\[ u_{k+1} = m^{k+1} \]

\[ m^{k+1} + a m^k + b m^{k-1} = 0 \]

\[ m^{k-1}(m^2 + a m + b) = 0 \]

Thus, we find $m$ by finding the roots of quadratic equation $x^2 + ax + b = 0$. This equation is also called the characteristic equation.

There are three solution scenarios for a quadratic equation:

1. 2 distinct real roots
2. 1 real root that is repeated
3. 2 complex roots

**Case 1:** Two distinct real roots $m_1$ and $m_2$, then, $u_k = m_1^k$ is one solution and $u_k = m_2^k$ is another solution. (This discussion assumes homogenous equations).
If you have two solutions, any linear combination of the two is also a solution. The general solution has the form:

\[ p_k = A m_1^k + B m_2^k \]

To review, the steps for solving a second order difference equation are:
1. Write out the homogeneous equation (divide through by the \( p_{k+1} \) coefficient to set it to 1)
2. Substitute the \( m \)’s
3. Write out the characteristic equation
4. Find roots
5. Substitute initial conditions to find \( A \) & \( B \)

**Case 2:** 1 real repeated root \( m \)

\[ u_k = m^k \] is a solution

\[ v_k = km^k \] is another solution

The general solution is \( p_k = A m^k + B km^k \)

**Gambler’s Ruin Problem**

A player plays a game and on each successive play of the game has the probability of \( p \) of winning $1 and probability of \( q \) of losing $1. If the player’s fortune reaches 0 or \( N \), the game is over.

Let \( p_k \) be the probability of losing if the player starts with \( k \) dollars.

\[ p_k = p \cdot p_{k+1} + q \cdot p_{k-1}; \ 0 < k < N \]

Initial conditions: \( p_0 = 1, \ p_N = 0 \)

Note that this is a second order linear equation with constant coefficients:

\[ p \cdot p_{k+1} - p_k + q \cdot p_{k-1} = 0 \]

\[ p_{k+1} - p_k/p + q/p \cdot p_{k-1} = 0 \]

\[ u_k = m^k \]

\[ x^2 - (1/p)x + q/p = 0 \]

---

1 Random walk is an example of a Markov chain with boundary conditions. Brownian motion, the classical example of a continuous stochastic process, is another example of a random walk that occurs as you take the limit as time step goes to 0.
m = 1, q/p

\[ p_k = A + B(q/p)^k \]

Substituting the initial conditions \( p_0 = 1, p_N = 0 \) yields:

\[ 1 = A + B \]
\[ 0 = A + B(q/p)^N \]

\[ p_k = ((q/p)^k - (q/p)^N)/(1 - (q/p)^N) \]

Take the limit as \( N \to \infty \).

Suppose that \( p < \frac{1}{2} \to q > \frac{1}{2}; q/p > 1; (q/p)^N \to \infty, N \to \infty \lim p_k = 1. \)

In case of a single root:

If \( m = 1, \ p_k = A + Bk \)

\[ 1 = A \]
\[ 0 = A + B/N \]

\[ -1 = B/N \]

\[ B = -1/N \]

\[ p_k = 1 - k/N \]

\[ N \to \infty \lim p_k = 1 \]

\[ p_k \to 1 \text{ as } N \to \infty, \text{ if } p \leq \frac{1}{2} \]