Probability 2 - Notes 2

Conditional probability

Definition If A and B are events and $P(B) > 0$ then we define the conditional probability of A given B to be

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

Definition Events $B_1, ..., B_n$ are said to partition the sample space S if $\cup_{i=1}^{n} B_i = S$ and $B_i \cap B_j = \phi$ for all $i \neq j$. (So the events are mutually exclusive and exhaustive)

The law of total probability

Let E be an event in S and let $B_1, ..., B_n$ partition S. Then

$$P(E) = \sum_{j=1}^{n} P(E|B_j)P(B_j)$$

You derived this result and looked at simple examples using this law in Probability 1

Example: A fair dice is thrown twice. Find $P(E)$ where E is the event that the product of the numbers on the die from the first and second throw is even. Let B_1 be the event that the number on the first throw is odd and B_2 be the event that the number on the first throw is even. then B_1, B_2 is a partition of S. $P(B_1) = P(B_2) = \frac{1}{2}$. If the number on the first throw is even then the product is certain to be even. So P$(E|B_1) = 1$. If the number on the first throw is odd, then the product will be even if the number on the second throw is even, so $P(E|B_2) = \frac{1}{2}$. Hence

$$P(E) = P(E|B_1)P(B_1) + P(E|B_2)P(B_2) = 1 \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{3}{4}$$

Use of the law of total probability for sequences of independent trials or games

Consider an independent sequence of throws of a die. We throw the die until the number on the die is either a 6 or is less than or equal to 2, when we stop. Let E be the event that we stop with a throw of 6. Find $P(E)$. A useful approach is to look back to the first throw of the die. Let B_1, B_2 and B_3 correspond to the event that the first throw gives respectively 6, less than or equal to 2 and neither 6 nor less than or equal to 2. Then B_1, B_2, B_3 is a partition.

$$P(E) = P(E|B_1)P(B_1) + P(E|B_2)P(B_2) + P(E|B_3)P(B_3)$$

Now $P(B_1) = \frac{1}{6}$, $P(B_2) = \frac{1}{3}$ and $P(E_3) = \frac{1}{2}$. Also $P(E|B_1) = 1$ and $P(E|B_2) = 0$. If B_3 occurs then after the first throw we are essentially in the same situation statistically as we were at the outset. We have a sequence of independent throws and will continue until the number thrown equals 6 or is less than 2. So $P(E|B_3) = P(E)$. Hence

$$P(E) = \frac{1}{6} + P(E)\frac{1}{2}$$

Therefore solving gives $P(E) = \frac{1}{3}$.

An application of this method is the gambler's ruin problem.

The gambler's ruin problem.

A gambler starts with a gambling pot or fund of k units of money. He plays a sequence of games. At each game he bets one unit and has a probability p of winning (in which case he receives 1 unit in addition to the 1 unit bet) and probability $q = 1 - p$ of losing (in which case he loses the 1 unit bet). He decides that he will stop if his pot/fund either grows to N units or declines to M units (in the original problem $M = 0$ so he goes broke i.e. is ruined). Let E_k be the event that when he stops he has reached N units and let $r_k = P(E_k)$.

We condition on the first outcome, so B_1 and B_2 are the events 'wins the first game' and 'loses the first game'. Then

$$r_k = P(E_k|B_1)P(B_1) + P(E_k|B_2)P(B_2)$$

If B_1 occurs then the gambler's stake has increased to $k+1$ and he is in the same situation as initially but with more stake money. Similarly if B_2 occurs he simply has less stake money ($k-1$ units). Hence $P(E_k|B_1) = P(E_{k+1}) = r_{k+1}$ and $P(E_k|B_2) = P(E_{k-1}) = r_{k-1}$. Therefore $r_k = pr_{k+1} + qr_{k-1}$. This is just a simple second order difference equation

$$pr_{k+1} - r_k + qr_{k-1} = 0$$

which holds for $M < k < N$. Note that $r_N = 1$ and $r_M = 0$ because the gambler stops immediately with a stake of M or N.

The associated quadratic is $p\theta^2 - \theta + q = 0$ which has roots $\theta = 1$ and $\theta = \frac{q}{p}$. The roots will be equal if $p = q = \frac{1}{2}$.

Case when $p \neq \frac{1}{2}$. The solution to the difference equation is

$$r_k = A(1)^k + B\left(\frac{q}{p}\right)^k = A + B\left(\frac{q}{p}\right)^k$$

Since $0 = r_M = A + B\left(\frac{q}{p}\right)^M$ and $1 = r_N = A + B\left(\frac{q}{p}\right)^N$, we obtain the solution

$$r_k = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^M}{\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^M}$$

Case $p = \frac{1}{2}$. The solution to the difference equation is $r_k = (A + Bk)(1)^k = A + Bk$. Since $0 = r_M = A + BM$ and $1 = r_N = A + BN$, we obtain the solution

$$r_k = \frac{k - M}{N - M}$$

Similarly if we let F_k be the event that the gambler has only M units when he stops playing and if $l_k = P(F_k)$. Then l_k satisfies the same difference equation as r_k but the boundary conditions are different since $l_M = 1$ and $l_N = 0$.

Case when $p \neq \frac{1}{2}$. The solution is

$$l_k = \frac{\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^k}{\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^M}$$

Note that $r_k + l_k = 1$ so the series of games are certain to finish.

Case $p = \frac{1}{2}$. The solution is

$$l_k = \frac{N - k}{N - M}$$

Again $r_k + l_k = 1$.

If we indicate in the notation the boundaries M and N then we replace r_k by $r_k(M,N)$ and l_k by $l_k(M,N)$ in the results above.

Note. The gambler's ruin problem is a special case of a random walk, which is a stochastic process. Here 'time' is the game number, so is discrete. For a random walk in discrete time the position at time n is Y_n where $Y_n = Y_{n-1} + X_n$. Here Y_{n-1} is the position at time $n-1$ and X_n is an independent increment. The X_j are i.i.d (independent identically distributed) random variables. If Y_0 is the starting position then $Y_n = Y_0 + \sum_{j=1}^{n} X_j$. There may be boundaries for the random walk (as in the gambler's ruin problem). The walk stops if the boundaries are reached.

The change in the gambler's stake after game j is a random variable X_j with $P(X_j = 1) = p$ and $P(X_j = -1) = q$. If Y_n is the amount he has immediately after the n^{th} game then $Y_n = k + \sum_{j=1}^{n} X_j$. This will only hold whilst the game is continuing. Once he reaches the boundary M or N the gambler stops playing. The 'stopping rule' implies that the number of games T played is a random variable. $T = \min\{j : \; Y_j = M \; or \; Y_j = N\}$.