Practical 2

Question 1

In a follow-up study of British doctors observations of smoking were made twice for each doctor, once in 1951 and once in 1991 (restricted to doctors still alive in 1991). The findings in 1951 were: Non-smoker 0.25, Ex-smoker 0.13 and Current smoker 0.62. The conditional distribution of the variable of smoking habits in 1991, given the outcome in 1951, is as follows:

		1991		
		Non-smoker	Ex-smoker	Current smoker
	Non-smoker	0.89	0.07	0.04
1951	Ex-smoker	0	0.95	0.05
	Current smoker	0	0.74	0.26

(a) Express this information in the form of a probability tree.

(b) What is the probability of being an Ex-smoker in 1991?

(c) Using Bayes' theorem obtain the conditional distribution of the smoking variable in 1951, given the doctor was an Ex-smoker in 1991.

Question 2

Cystic fibrosis is an inherited condition in which the lungs and digestive system can become clogged with thick, sticky mucus. The gold standard diagnostic test is the sweat chloride test. The sensitivity of the sweat chloride test is 91.7% and the specificity is 99.9%.

Tests are only performed on patients suspected of having cystic fibrosis (i.e. those with symptoms typical of cystic fibrosis). The probability of cystic fibrosis in this group is 3.25%.

(a) Calculate the positive predictive value for this test.

(b) Interpret the probability you calculated in (a).

(c) In the whole population, the probability of cystic fibrosis is 0.04%. Would this test be useful as a population screening test?

Question 3

The table below shows the joint distribution of two discrete random variables X and Y. For each value of x and y, the table shows $P(X = x, Y = y)$.

		y: 1	2	3	Total
	0	0.05	0.1	0.15	0.3
x	1	0	0.05	0.05	0.1
	2	0.05	0.35	0.2	0.6
	Total	0.1	0.5	0.4	1

(a) Calculate $E(X)$ and $Var(X)$.

(b) Using the information in the table, determine if X and Y are independent.

(c) The conditional expectation $E(Y|X = x)$ is defined as the expectation (in the usual way) with respect to the conditional distribution $P(Y|X = x)$.

(i) Find the conditional distributions $P(Y|X = 0)$.

(ii) Calculate $E(Y|X = 0)$.

$P(Y = 1|X = 1) = 0$, $P(Y = 2|X = 1) = 0.5$, and $P(Y = 3|X = 1) = 0.5$, and the conditional expectation given $X = 1$ is $E(Y|X = 1) = 2.5$.

(iii) Do your answers to (i) and (ii) make sense in light of your answer to (b)?

Additional: Question 4

Let X be a discrete random variable with $E(X) = 1$ and $Var(X) = 5$. Find

(a) $E(2 + 3X)$

(b) $E((2 + X)^2)$

(c) $Var(10 + 3X)$

Additional: Question 5

If X and Y are independent discrete random variables, prove that:

(a) $E(XY) = E(X)E(Y)$.

(b) $Var(X + Y) = Var(X) + Var(Y)$

Optional: Brainteaser

Two MSc Medical Statistics students, A and B, play a game in which they each throw a ball and the winner is the one who throws the furthest. Their throws are independent and the two students are equally skillful, i.e. it is a game of pure chance.

Supposing that A and B get one throw each, we would all agree (I hope) that B's probability of winning is $\frac{1}{2}$.

Now suppose that A throws once, but that B gets two throws, the winner being the thrower of the ball which goes the furthest. What now is the probability that B wins?

Here are two potential solutions to this question:

Solution 1

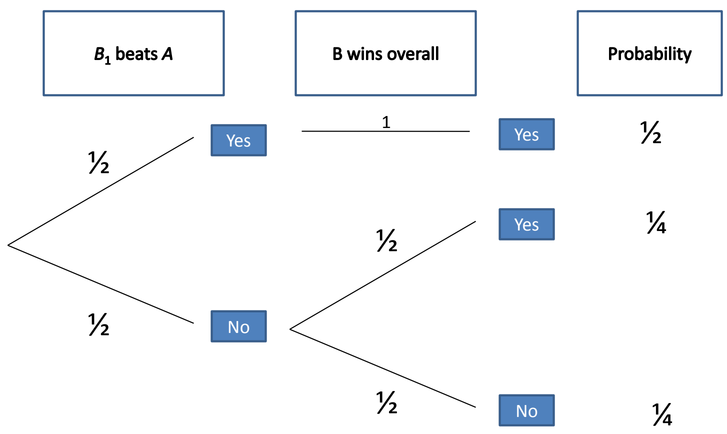
Let A be the length of A's throw and B_1 and B_2 be the lengths of B's two throws. If we write A, B_1 and B_2 in increasing order of magnitude, there are 6 possible permutations:

$$(A, B_1, B_2), (A, B_2, B_1), (B_1, A, B_2), (B_1, B_2, A), (B_2, A, B_1), (B_2, B_1, A)$$

Since both students are equally skillful, these 6 permutations occur with equal probability. A can only win if A's throw is the furthest of the three, which happens in 2 out of the 6 permutations. Thus, A's probability of winning is $\frac{1}{3}$ and **B's probability of winning is $\mathbf{\frac{2}{3}}$**.

Solution 2

Suppose that A throws first. Now, B throws his first ball and has probability $\frac{1}{2}$ of beating A with his first throw. If he manages this, then he has won the game no matter what happens to his second throw. However, if loses, he then has a second chance to beat A's throw. His first throw is now irrelevant, and his second throw beats A's throw with probability $\frac{1}{2}$. This can be illustrated using the following tree diagram:



Thus, **B wins with probability $\mathbf{\frac{3}{4}}$**.

Which of these solutions is correct? And—more importantly—why is the other incorrect?