

Problem 72: Commuting by train

(✓✓)

Tabulated values of $\Phi(\cdot)$, the cumulative distribution function of a standard normal variable, should not be used in this question.

Henry the commuter lives in Cambridge and his working day starts at his office in London at 0900. He catches the 0715 train to King's Cross with probability p , or the 0720 to Liverpool Street with probability $1 - p$. Measured in minutes, journey times for the first train are $N(55, 25)$ and for the second are $N(65, 16)$. Journey times from King's Cross and Liverpool Street to his office are $N(30, 144)$ and $N(25, 9)$, respectively. Show that Henry is more likely to be late for work if he catches the first train.

Henry makes M journeys, where M is large. Writing A for $1 - \Phi(\frac{20}{13})$ and B for $1 - \Phi(2)$, find, in terms of A , B , M and p , the expected number, L , of times that Henry will be late and show that, for all possible values of p ,

$$BM \leq L \leq AM.$$

Henry noted that in $\frac{3}{5}$ of the occasions when he was late, he had caught the King's Cross train. Obtain an estimate of p in terms of A and B .

Note: A random variable is said to be $N(\mu, \sigma^2)$ if it has a normal distribution with mean μ and variance σ^2 .

2000 Paper II

Comments

This is impossible unless you know the following result:

If X_1 and X_2 are independent and normally distributed according to $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, then $X_1 + X_2$ is also normally distributed and $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Even if you didn't know this result, you do now and it should not be difficult to complete the first parts of the question.

For the last part, you need to know something about conditional probability, namely that

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

which makes sense intuitively and can easily be understood in terms of Venn diagrams. The denominator is essentially a normalising constant. The formula may be taken as the definition of conditional probability on the left-hand side.

Solution to problem 72

Let T_1 be the random variable representing the total journey time via Kings Cross, so that

$$T_1 \sim N(55 + 30, 25 + 144) = N(85, 169),$$

using the result mentioned on the previous page, and let T_2 be the random variable representing the total journey time via Liverpool Street, so that

$$T_2 \sim N(65 + 25, 16 + 9) = N(90, 25).$$

Then the probabilities of being late are, respectively, $P(T_1 > 105)$ and $P(T_2 > 100)$, i.e. $1 - \Phi(\frac{20}{13})$ and $1 - \Phi(2)$. Note that $\Phi(\frac{20}{13}) < \Phi(2)$.

We have

$$L = [pA + (1 - p)B]M = [B + (A - B)p]M.$$

L increases as p increases, since $A > B$, hence the given inequalities corresponding to $p = 0$ and $p = 1$.

We have

$$P(\text{Kings Cross given late}) = \frac{P(\text{Late and Kings Cross})}{P(\text{Late})}$$

An estimate for p (call it $\tilde{p}$) is therefore given by

$$\frac{3}{5} = \frac{A\tilde{p}}{A\tilde{p} + B(1 - \tilde{p})},$$

$$\text{so } \tilde{p} = \frac{3B}{2A + 3B}.$$

Post-mortem

The result mentioned in the comment on the previous page is just the sort of thing you should try to prove yourself rather than take on trust. Unfortunately, such results in probability tend to be pretty hard to prove. You can also prove it fairly easily using generating functions (which are not on the STEP specifications any longer). You can also prove it from first principles. That would be difficult because of the awkward integrals involved, but not impossible.