STEP II, 2022, Q11

11 A batch of N USB sticks is to be used on a network. Each stick has the same unknown probability p of being infected with a virus. Each stick is infected, or not, independently of the others.

The network manager decides on an integer value of T with $0 \leqslant T < N$. If $T = 0$ no testing takes place and the N sticks are used on the network, but if $T > 0$, the batch is subject to the following procedure.

- Each of T sticks, chosen at random from the batch, undergoes a test during which it is destroyed.
- If any of these T sticks is infected, all the remaining $N - T$ sticks are destroyed.
- If none of the T sticks is infected, the remaining $N - T$ sticks are used on the network.

If any stick used on the network is infected, the network has to be disinfected at a cost of £D, where $D > 0$. If no stick used on the network is infected, there is a gain of £1 for each of the $N - T$ sticks. There is no cost to testing or destroying a stick.

(i) Find an expression in terms of N, T, D and q, where $q = 1 - p$, for the expected net loss.

(ii) Let $\alpha = \dfrac{DT}{N(N - T + D)}$. Show that $0 \leqslant \alpha < 1$.

Show that, for fixed values of N, D and T, the greatest value of the expected net loss occurs when q satisfies the equation $q^{N-T} = \alpha$.

Show further that this greatest value is £$\dfrac{D(N-T)\,\alpha^k}{N}$, where $k = \dfrac{T}{N-T}$.

(iii) For fixed values of N and D, show that there is some $\beta > 0$ so that for all $p < \beta$, the expression for the expected loss found in part **(i)** is an increasing function of T. Deduce that, for small enough values of p, testing no sticks minimises the expected net loss.