J09T.3 - Errors in Gene Expression

Problem

A biased coin has a probability p of coming up heads in a single toss.

a) Write down the probability P_n of obtaining n heads in N tosses.

b) Show that, in the limit $p \ll 1$ and $n \ll N$, P_n reduces to the Poisson formula

$$P_n = \frac{\lambda^n}{n!} e^{-\lambda}, \qquad (\lambda \equiv Np).$$

[Hint: Consider $\ln P_n$. Stirling's approximation is $n! \simeq n^n e^{-n}$. Note that $\lambda \gg p$.]

c) Assume P_n has the Poisson form, and write $\sum_{n=0}^{\infty} n^s P_n = \langle n^s \rangle$. Calculate the mean $\bar{n} = \langle n \rangle$, and the variance $\langle (n - \bar{n})^2 \rangle$.

d) A single strand of DNA is comprised of a string of basic units (the nucleotides A, T, G, and C) arrayed in a specific order. During gene expression, a molecular machine crawls along the strand and reads the units sequentially (this starts a chemical sequence that assembles one protein molecule).

The machine is highly reliable but not infallible. On average, it makes only one error for every 10^6 units read. Assume that a gene is comprised of $N_{\text{gene}} = 3 \times 10^4$ units. Calculate the probability that the machine makes zero errors when the gene is read. Find the probability that it makes 2 errors.