CS 70 Challenge Problems: Expectation, Variance, and Bounds

Solutions at https://alextseng.net/teaching/cs70/

Alex Tseng

1 Expectation

(a) Short tandem repeats (or STRs) occur in some segments of DNA. They consist of very short sequences that are repeated usually around 20-30 times. Consider a strand of DNA with n base pairs, where each base could be **A**, **C**, **G**, or **T**, randomly and independently. We are interested in the STR **AG**, so we are looking for segments of DNA that look like **AGAGAG**. This segment would have 3 repeats. What is the probability that you have at least r repeats starting at position p? Assume that all $2r$ bases would fit in the n base pairs starting at p.

(b) Continuing from the above part, what is the expected number of positions where a run of at least r repeats could start?

(c) You draw a random number from the set $\{1, ..., 100\}$, and then another number from the set $\{1, ..., 50\}$. What is the expectation of the sum of the numbers? What is the expectation of the product?

(d) *Challenge* Suppose you have chips numbered 1 to k in a bag, and you draw 2 from the bag randomly, *with replacement*. Let A be the first number you drew, and B be the second number you drew. What is $\mathrm{E}[\max\{A, B\}]$? What is $\mathrm{E}[\min\{A, B\}]$. Show that $\mathrm{E}[\max\{A, B\}] + \mathrm{E}[\min\{A, B\}] = \mathrm{E}[A] + \mathrm{E}[B]$ *without appealing to the linearity of expectations.*
Hint: Start with the definition of expected value. You will need to be a little clever with the algebra.

2 Variance

(a) Imagine we roll a standard 6-sided die 100 times, and add up the resulting values. What is the variance of this distribution?

(b) Prove or give a counterexample: if X and Y are independent, then $\mathrm{Var}[X - Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

3 Bounds

(a) If the national IQ average is 100, and $\sigma = 10$ (recall σ is standard deviation), what is the probability of finding someone with an IQ of 300 or more? Give the tightest bound you can.

(b) Consider a distribution X with $\mu = 13, \sigma = 5$. We know that X never exceeds 17. Is it possible that $P(X \leq 1) > \frac{1}{4}$?