

$X \sim \text{Bin}(n, p)$      $n$  indep coin tosses with prob  $p$  of H

$X$  counts # H's. (or successes)

$$E(X) = np$$

$$\text{Var}(X) = E(X^2) - \mu^2$$

$$E(X^2) = E[(X_1 + X_2 + \dots + X_n)(X_1 + X_2 + \dots + X_n)]$$

$$\begin{aligned} &= \sum_{i=1}^n \underbrace{E(X_i^2)}_p + \sum_{i=1}^n \sum_{j \neq i}^n \underbrace{E(X_i X_j)}_{\substack{= \Pr(X_i=1, X_j=1) \\ \text{indep} \\ = p^2}} \end{aligned} \quad \text{linearity of expectation}$$

$$= np + n(n-1)p^2$$

$$\text{Var}(X) = E(X^2) - \mu^2 = np + n(n-1)p^2 - n^2p^2$$

$$= np(1-p)$$

(for which  $p$  is it maximized?)

DNA sequence each position indep A, G, T, C

with prob  $P_A$   $P_G$   $P_T$   $P_C$

length = 10,000,000

Expected # of occurrences of AATGAAT?

AATGAATGAATCC