Problem 1
In your 1st project you will be working with the HP model discussed in class. What is the maximum amount of contacts that a single residue in a chain can make?
What is the lowest possible energy that we can obtain from a sequence. What parameters of the sequence are important for this bound?
Draw a conformation on the lattice with energy \( \leq -3 \) for the sequence:

\[ HPHPHPHPPH \]

Using the same conformation as you did above, show how you can change the energy of the conformation by exchanging 2 of the residues' locations (e.g. exchange locations of residue 1 and 5).
Using the original sequence above, draw a different conformation on the lattice with energy \( \leq -3 \).

Solution 1
A single residue can make up to 3 contacts (an endpoint surrounded from all directions).
I accepted any bound on the lowest energy. A trivial bound would be \(-3\) times the length of the sequence. A tighter bound would be \(-1\) the number of H's in the sequence \(-1\). There can be even tighter bounds.
Reasonable parameters for such lower bound on the energy are sequence length, number of H's in the sequence, number of H's in the sequence of even (or odd) parity, etc.
All of you had multiple good conformations with energy of -3 or less and shown how change of sequence can change a structure's energy.
Problem 2
Assume that a computer can generate truly uniformly distributed numbers on the interval (0, 1). Let $X_1$ be the first number that the computer generates, $X_2$ be the second such number generated and so on ($X_n$ will of course be the $n^{th}$ such number generated).

What is the probability that the first number generated is more than 0.2 away from the expected value (i.e. $P(|X_1 - 0.5| \geq 0.2)$? What bounds can Chebyshev’s inequality provide here?

What is the probability that the average of the first two numbers generated is more than 0.2 away from the expected value (i.e. $P(|\frac{X_1 + X_2}{2} - 0.5| \geq 0.2)$? What bounds can Chebyshev’s inequality provide in this case?

Using Chebyshev’s inequality, give a bound to the probability that the average of the first ten numbers generated is more than 0.2 away from the expected value (i.e. for $P(|\frac{\sum_{i=1}^{10} X_i}{10} - 0.5| \geq 0.2)$.

How many numbers generated by the computer should you average in order to get that the probability of this average being more than 0.2 away from the mean will be less than 0.1% (i.e. $P(|\overline{X_n} - 0.5| \geq 0.2) < 0.1\%$)?

Solution 2

$P(|X_1 - 0.5| \geq 0.2) = P((X_1 - 0.5) \geq 0.2 \text{ or } (0.5 - X_1) \geq 0.2) = P(X_1 \geq 0.7 \text{ or } X_1 \leq 0.3) = P(X_1 \geq 0.7) + (X_1 \leq 0.3) = (1 - \frac{7}{10}) + \frac{3}{10} = \frac{6}{10}$

Using Chebyshev inequality $P((X - \mu)^2 \geq t^2) = P(|X - \mu| \geq t) \leq \frac{\mu^2}{t^2}$ with $\mu = 0.5$ and $\sigma = \frac{1}{12}$ (see previous homework) we get the approximations:

$P(|X_1 - 0.5| \geq 0.2) \leq \frac{\frac{6}{10^2}}{12^2} = \frac{25}{12^2} > 1$

From homework 1 we have the p.d.f. of a random variable $W = X_1 + X_2$ given as $F_W(w) = \begin{cases} 0 & w < 0 \\ \frac{w^2}{2} & 1 \leq w \leq 0 \\ 2w - \frac{w^2}{2} - 1 & 1 \leq w \leq 2 \\ 1 & 2 < w \end{cases}$

We will use this knowledge in our solution.

$P(|\frac{X_1 + X_2}{2} - 0.5| \geq 0.2) = P(|X_1 + X_2 - 1| \geq 0.4) = P(|W - 1| \geq 0.4) = P((W - 1) \geq 0.4 \text{ or } (1 - W) \geq 0.4) = P(W \geq 1.4 \text{ or } W \leq 0.6) = 1 - P(0.6 \leq W \leq 1.4) = 1 - (P(1.4 \leq W) - P(0.6 \leq W)) = 1 - (\frac{1.4^2}{2} - (2*0.6 - 0.6^2 - 1)) = 1 - (\frac{1.96}{2} - 1.2 + \frac{0.36}{2} + 1) = 1.2 - 0.98 + 0.18 = 0.36$

To find the Chebyshev bound all we need to do is to find the new mean and variance.

$\mu = E(\frac{X_1 + X_2}{2}) = \frac{1}{2}E(X_1 + X_2) = \frac{1}{2}(E(X_1) + E(X_2)) = \frac{1}{2} \cdot 2 \cdot E(X) = E(X) = 0.5$

$\sigma = Var(\frac{X_1 + X_2}{2}) = \frac{1}{4}Var(X_1 + X_2) = \frac{1}{4}(Var(X_1) + Var(X_2)) = \frac{1}{4} \cdot 2 \cdot (\frac{1}{12} \cdot \frac{1}{2}) = \frac{1}{12}$

We now use Chebyshev inequality and compute $P(|X - \mu| \geq t) \leq \frac{\mu^2}{t^2}$.
\[ P\left( \frac{X_1 + X_2}{2} - 0.5 \right) \geq 0.2 \leq \frac{1}{24+0.04} = \frac{25}{24} > 1 \]

Similarly for \( \overline{X}_{10} = \frac{\sum_{i=1}^{10} X_i}{10} \) the mean and variance will be:
\[ E(\overline{X}_{10}) = \frac{1}{10} \sum_{i=1}^{10} E(X_i) = \frac{1}{10}(10E(X)) = E(X) = 0.5 \]
\[ Var(\overline{X}_{10}) = \frac{1}{100} \sum_{i=1}^{10} Var(X_i) = \frac{1}{100}(10Var(X)) = \frac{1}{10} Var(X) = \frac{1}{10} \frac{1}{4} = \frac{1}{40} \]
Again, use Chebyshev inequality and compute \( P(\sqrt{40} \geq t) \leq \frac{2}{t^2} \):
\[ P(\overline{X}_{10} - 0.5 \geq 0.2) \leq \frac{1}{120+0.04} = \frac{25}{129} \approx 0.2083 \]

Now we reverse engineer and calculate mean and variance for \( \overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \).
\[ E(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n}(nE(X)) = E(X) = 0.5 \]
\[ Var(\overline{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}(nVar(X)) = \frac{1}{n} Var(X) = \frac{1}{n} \frac{1}{12} = \frac{1}{12n} \]
We use Chebyshev inequality and see:
\[ P(\overline{X}_n - 0.5 \geq 0.2) \leq \frac{1}{12n+0.04} = \frac{25}{12n} < 0.1\% = 0.001 \]
So \( \frac{25}{12} < 0.001 \Rightarrow \frac{25000}{12} \leq n \Rightarrow n \geq 2084. \)