DIS 13A

1 Chebyshev's Inequality vs. Central Limit Theorem

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the following distribution:

$$\mathbb{P}[X_1 = -1] = \frac{1}{12}; \qquad \mathbb{P}[X_1 = 1] = \frac{9}{12}; \qquad \mathbb{P}[X_1 = 2] = \frac{2}{12}.$$

(a) Calculate the expectations and variances of X_1, $\sum_{i=1}^n X_i$, $\sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, and

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}.$$

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n| \geq 2]$.

(c) Can you use b to bound $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$?

(d) As $n \to \infty$, what is the distribution of Z_n?

(e) We know that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{P}[|Z| \leq 2] = \Phi(2) - \Phi(-2) \approx 0.9545$. As $n \to \infty$, can you provide approximations for $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$?

2 Binomial Concentration

Here, we will prove that the binomial distribution is *concentrated* about its mean as the number of trials tends to ∞. Suppose we have i.i.d. trials, each with a probability of success $1/2$. Let S_n be the number of successes in the first n trials (n is a positive integer), and define

$$Z_n := \frac{S_n - n/2}{\sqrt{n}/2}.$$

(a) What are the mean and variance of Z_n?

(b) What is the distribution of Z_n as $n \to \infty$?

(c) Use the bound $\mathbb{P}[Z > z] \leq (\sqrt{2\pi}z)^{-1}\,\mathrm{e}^{-z^2/2}$ when Z is normally distributed in order to bound $\mathbb{P}[S_n/n > 1/2 + \delta]$, where $\delta > 0$.

3 Correlation and Independence

(a) What does it mean for two random variables to be uncorrelated?

(b) What does it mean for two random variables to be independent?

(c) Are all uncorrelated variables independent? Are all independent variables uncorrelated? If your answer is yes, justify your answer; if your answer is no, give a counterexample.

4 Covariance

We have a bag of 5 red and 5 blue balls. We take two balls from the bag without replacement. Let X_1 and X_2 be indicator random variables for the first and second ball being red. What is $\operatorname{cov}(X_1, X_2)$? Recall that $\operatorname{cov}(X_1, X_2) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$.