Assignment 8

CS2244-Randomness and Computation
Assigned April 16
Due April 26

April 16, 2018

These problems are mostly about the Law of Large Numbers and the Central Limit Theorem.

1 Marbles

A jar contains billions of marbles. You know that 48% of them are black and the rest white. You pull out 1000 marbles. What is the probability that more than half the marbles you draw are black? Use the normal approximation to the binomial distribution to compute this.

2 Dice

One of the spots on the 3 face of a die has fallen off, thus changing the 3 to a 2. Let X denote the outcome (an element of the set $\{1, 2, 4, 5, 6\}$) of a roll of this die.
(a) Calculate and plot the PMF of X.

(b) Calculate $E(X)$.

(c) Calculate $Var(X)$ and σ_X (the standard deviation).

(d) Use Chebyshev's inequality to find a lower bound on the probability $P[3 \leq X \leq 4]$. Note that I am not asking you to compute this probability, which is easy to do. Instead I am asking you to see what kind of estimate the inequality gives in this case. Note that Chebyshev's inequality in its basic form gives an *upper*

bound on the probability that a random variable deviates at least a certain amount from its mean. Here you are asked about a lower bound on the complementary probability.

Now consider n successive rolls of this damaged die. Let X_i denote the outcome of the i^{th} roll and S_n the sum $X_1 + \cdots + X_n$.

(e) Use Chebyshev's inequality to find a lower bound on the probability $P[3n \leq S_n \leq 4n]$. Along the way, you will have to compute $E(S_n)$, $Var(S_n)$ and σ_{S_n}.

(f) The Central Limit Theorem should apply here. Use it to estimate the probability $P[150 \leq S_{50} \leq 200]$. (Remember that since S_{50} is a discrete random variable, this is the same as the probability $P[149 < S_{50} < 201]$, so you will get slightly different estimates depending on the exact interval you use to make the estimate.)

(g) Now simulate 50 rolls of the altered die, repeating the experiment a large number of times (try between 1000 and 10000 repetitions). Use this to estimate the probability in (f). The answer should be close to the normal approximation.

(h) Use the simulation results to create a plot of the PMF of S_{50} normalized so that it has mean 0 and standard deviation 1, and superimpose a plot of the standard normal density on it.

(i) Another way to display the near-normality of S_{50} is to use the 'probability plot' provided by `probplot`. (I will explain probability plots and quantile plots in class.)To use this, you need to first

```
import scipy.stats
```

and then call

```
v=scipy.stats.probplot(x,plot=plt)
plt.show()
```

where x is a list of the simulation results. Normal distributions should show values lying along a straight line.

3 A Question About Limits

This problem is adapted from the Grinstead and Snell book, and is a good test of your understanding of how the Law of Large Numbers and the normal distribution work. A coin with probability 0.8 of heads is tossed n times. S_n denotes the number of heads and $A_n = S_n/n$ the average number of heads. The problem is to evaluate each of the following limits. (There are only a couple parts where you have to do any computation.)

(a) $\lim_{n\to\infty} P(A_n = 0.8)$

(b) $\lim_{n\to\infty} P(0.7n < S_n < 0.8n)$

(c) $\lim_{n\to\infty} P(S_n < 0.8n + \sqrt{n})$

(d) $\lim_{n\to\infty} P(0.79 < A_n < 0.81)$

(e) $\lim_{n\to\infty} P(0.79 < A_n < 0.8 + \frac{1}{\sqrt{n}})$