DS 100: Principles and Techniques of Data Science **Date: January 26th, 2017**

Discussion #1 Solutions

1. **Probability & Combinatorics**

(a) In the Powerball Lottery, there are $N = 69$ balls in a cage, numbered 1 through 69, and in the first part of the lottery, $r = 5$ are sampled without replacement by repeatedly sampling a ball uniformly from the remaining subset. How many subsets of five balls could we possibly draw?

answer: $\binom{N}{r} = \binom{69}{5} = 11,238,513$ subsets. See below for an explanation.

You get N choices for the first ball, $N-1$ choices for the second ball, and so on. If you are drawing r balls without replacement, then the number of sequences of length r without repetitions is

$$\underbrace{N(N-1)(N-2)\cdots(N-r+1)}_{r \text{ terms}} = \frac{N!}{(N-r)!}.$$

But since we don't care about the order of the r balls, for each sequence above we overcounted by a factor of $r!$, so the number of subsets is $\frac{N!}{(N-r)!\,r!} = \binom{N}{r}$.

(b) If you correctly pick 3 of the 5 numbers drawn in the Powerball Lottery, you win \$8 (a ticket costs \$2). What is the chance of correctly picking 3 of the 5 numbers drawn, assuming you guess the numbers at random?

answer: $\frac{\binom{5}{3}\cdot\binom{64}{2}}{\binom{69}{5}} = \frac{6720}{3{,}746{,}171} \approx 1.8\%$. See below for an explanation.

Since we are guessing at random, we are looking for the number of subsets containing exactly 3 of the 5 numbers drawn, divided by the total number of subsets from part (a). There are $\binom{5}{3}$ ways to pick 3 numbers from the correct 5, and $\binom{64}{2}$ ways to choose the remaining 2 numbers from the remaining incorrect 64.

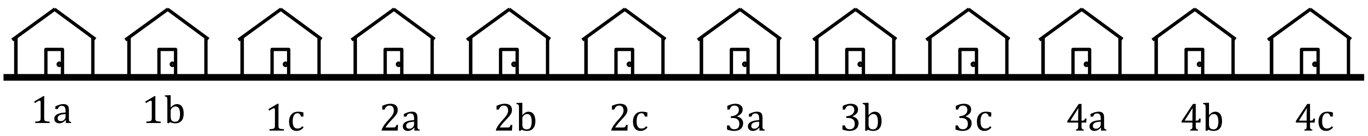
(c) To win the Jackpot in the Powerball Lottery, one must choose the correct subset of 5 balls from part (a), as well as matching a 'megaball' chosen uniformly between 1 and 35 from a different cage. What is the chance of winning the Jackpot, assuming you guess the numbers at random?

answer: Multiply the probabilities of each event: $\frac{1}{\binom{69}{5}\cdot 35} \approx 2.5 \cdot 10^{-9}$.

(d) What is the probability that the second regular ball is numbered 20?

answer: $\frac{1}{69}$ (it doesn't matter that we're considering the second ball drawn since we don't know anything about the result of the first).

2. **Sampling Strategies**



Kalie wants to measure interest for a party on her street. She assigns numbers and letters to each house on her street as illustrated above. She picks a letter “a”, “b”, or “c” at random and then surveys every household on the street ending in that letter.

(a) What kind of sample has Kalie collected? *answer:* a cluster sample (each group of houses ending in a certain letter is a cluster).

(b) What is the chance that two houses next door to each other are both in the sample? *answer:* None of the adjacent houses end in the same letter, so the chance is zero.

(c) Now suppose Kalie instead picks one house beginning with ‘1’ at random, one house beginning with ‘2’ at random, and so on, so she surveys four houses, one of each number. What kind of sample has Kalie collected? *answer:* a stratified sample.