Discussion 6: Week 3 - Stats 100 A

Andrew Lizarraga

Department of Statistics and Data Science

June 27, 2024

Initial Remarks

- Homework 1 is graded.

Initial Remarks

- Homework 1 is graded.
- On the google form, most of you indicated problems 4 and 5 being the most difficult for Homework 2.

Problem 4 (1 - 2)

Problem 4 Suppose a population has N people. Among them, M are males. Suppose we randomly sample a person uniformly from the population.

(1) What is the probability p of getting a male?

(2) If we repeat the above uniform random sampling from the same population n times independently. Let Ω_n be the sample space of all possible sequences. These sequences are all equally likely by definition of (a) uniform sampling and (b) independence. What is the number of sequences in Ω_n?

Problem 4 (3)

(3) For a sequence $\omega \in \Omega_n$, let $X(\omega)$ be the number of males in this sequence. For each $x \in \{0, 1, ..., n\}$, define $A_x = \{\omega : X(\omega) = x\}$ be the set of sequences with exactly x males. What is the number of sequences in A_x? Calculate $p(x) = P(A_x) = P(X = x)$ using Axiom 0. Explain that $p(x)$ can also be used as the probability distribution of the number of heads in coin flipping where p is the probability of head in each flip.

Problem 4 (4)

(4) Let

$$A = \left\{ \omega : \left| \frac{X(\omega)}{n} - p \right| \le .01 \right\}$$

be the set of representative sequences whose frequencies are close to p. Calculate the number of sequences in A. Calculate $P(A)$ using Axiom 0.

Hint: we can re-express the absolute term as follows:

$$p + (0.01) \le \frac{X(\omega)}{n} \le p + (0.01)$$

$$np + n(0.01) \le X(\omega) \le np + n(0.01)$$

$$\lceil np + n(0.01) \rceil \le X(\omega) \le \lfloor np + n(0.01) \rfloor$$

Thus

$$|A| = \sum_{x = \lceil np + n(0.01) \rceil}^{\lfloor np + n(0.01) \rfloor} \binom{n}{x}$$

Therefore

$$P(A) = \frac{|A|}{\Omega_n} = \frac{\sum_{x = \lceil np + n(0.01) \rceil}^{\lfloor np + n(0.01) \rfloor} \binom{n}{x}}{2^n}$$

Problem 5 (1)

Problem 5 Consider a random walk over 3 webpages, 1, 2, 3. At any step, if the person is at webpage 1, then with probability 1/6, she will go to webpage 2, and with probability 1/6, she will go to webpage 3. If the person is at webpage 2, then with probability 1/2, she will go to webpage 1, and with probability 1/2, she will go to webpage 3. If the person is at webpage 3, then with probability 1/2, she will go to webpage 1, and with probability 1/2, she will go to webpage 2.

Let X_t be the webpage the person is browsing at time t, and let us assume she starts from webpage 1 at time 0, i.e., $X_0 = 1$.

(1) Let $K_{ij} = P(X_{t+1} = j | X_t = i)$. Let $K = (K_{ij})$ be the 3×3 transition matrix. Write down K.

Problem 5 (2)

(2) Let $p_i^{(t)} = P(X_t = i)$. Let $p^{(t)} = (p_i^{(t)}, i = 1, 2, 3)$ be the row vector. Calculate $p^{(t)}$ for $t = 1, 2, 3$ using vector matrix multiplication.

Problem 5 (3)

(3) Let π_i be the stationary distribution at webpage i, so that $\pi_j = \sum_{i=1}^{3} \pi_i K_{ij}$, and $\sum_{i=1}^{3} \pi_i = 1$. Given the above K, solve π_i, $i = 1, 2, 3$, from these equations. Let $\pi = (\pi_i, i = 1, 2, 3)$ be the row vector. Is $p^{(3)}$ close to π?

Problem 5 (3)

(4) Based on the above calculations, answer the following questions. Suppose there are 1 million people doing the above random walk independently, and suppose they all start from webpage 1 at time $t = 0$. Then what is the distribution of these 1 million people for $t = 1, 2, 3$? What is the stationary distribution of these 1 million people? Which page is the most popular?

Thank You / Questions

Contact:

You may contact me at: **andrewlizarraga@g.ucla.edu**.