Stat 201A, Fall 2024: Lab 4

Conceptual review

- If X and Y are independent continuous random variables with probability density functions f and g. What is the probability density function of $X + Y$?
- What is the statement of CLT for Binomail distribution?
- What does the Kullback–Leibler divergence describes?

Problem 1

1. Roll a fair die 720 times. Estimate the probability that we have exactly 113 sixes.
2. You flip a fair coin 10,000 times. Approximate the probability that the difference between the number of heads and number of tails is at most 100.

Problem 2

Suppose we have a biased coin and we do not know the true probability p that it lands on heads. How can we estimate p? Can we estimate the error of our approximation?

Problem 3

Mitchell and Alex are competing together in a 2-mile relay race. The time Mitchell takes to finish (in hours) is $X \sim \text{Unif}(0, 2)$ and the time Alex takes to finish his mile (in hours) is continuous $Y \sim \text{Unif}(0, 1)$. Alex starts immediately after Mitchell finishes his mile, and their performances are independent. What is the distribution of $Z = X + Y$, the total time they take to finish the race?

Problem 4

Suppose $p_1, \ldots, p_k$ be a set of nonnegative numbers that sum to one. Suppose $f_1, \ldots, f_k$ are another set of nonnegative numbers that sum to one. The Kullback-Leibler divergence between these two sets of numbers is given by

$$KL(f\|p) := \sum_{i=1}^{k} f_i \log \frac{f_i}{p_i}$$

1. Show that $KL(f\|p)$ is always nonnegative.
2. Show that $KL(f\|p) = 0$ if and only if $f_i = p_i$ for each $i = 1, \ldots, k$.
3. Suppose that a coin toss can give three different results: H (heads), T (tails) and edge (when the coin just stands on its edge). Suppose that a person A assigns probabilities

   $$p_1^A = 0.499, \quad p_2^A = 0.499, \quad p_3^A = 0.002$$

   to the three outcomes and another person B assigns probabilities

   $$p_1^B = p_2^B = p_3^B = \frac{1}{3}$$

   to the three outcomes. Suppose that an experiment is performed by tossing the coin a bunch of times and this led to the observed proportions

   $$f_1 = \frac{14}{29}, \quad f_2 = \frac{14}{29}, \quad f_3 = \frac{1}{29}$$

   of the three outcomes. Calculate the Kullback-Leibler divergences $KL\left(f\|p^A\right)$ and $KL\left(f\|p^B\right)$. Which of $KL\left(f\|p^A\right)$ and $KL\left(f\|p^B\right)$ is smaller and does that seem reasonable?

Problem 5

Compare real binomial probabilities with entropy and normal approximations for $n = 100$ and $p = 0.5$ and $p = 0.05$, using plots to visualize. You may play around with different n, p and k to compare.