# Foundations of Data Science & Machine Learning

Tutotial 06

April 9, 2021

**Question 1.** There are some red and blue balls in a jar. Your random experiment involves sampling 100 balls out of it uniformly and independently (with repetition). You repeat this experiment twice and count the number of red balls you get in each repetition, say $R_1$ and $R_2$. Show that with probability at least 0.45, $|R_1 - R_2| \leq 20$.

   *Hint.* Hoeffding Bounds.

**Hoeffding Bounds.** If $X$ is the average of $n$ $\{0, 1\}$-random variables $X_1, \ldots, X_n$ with $P(X_i = 1) = p$ for each $i \in [n]$, then

$$P(X > p + \epsilon) \leq e^{-2\epsilon^2 n}, \text{ and}$$
$$P(X < p - \epsilon) \leq e^{-2\epsilon^2 n}.$$

**Question 2.** Suppose you know that there is a total of 1000 balls in the jar and $R_1$ was 30 in the above experiment. You want to estimate $R$, the total number of red balls in the jar. What is the smallest interval $[a, b]$ for which you can guarantee that $R \in [a, b]$ with probability at least 0.9?