Machine Learning A (2023)
Home Assignment 2

Niels Moctezuma Krarup. Student ID: WTG176

Contents

1 Illustration of Markov's, Chebyshev's, and Hoeffding's Inequalities (24 points)

2a

subsection 2a

1

We start by plotting the empirical distribution based on 1e6 replications of observing the mean being above $\alpha \in \{0.5, 0.55, \ldots, 0.95, 1\}$.

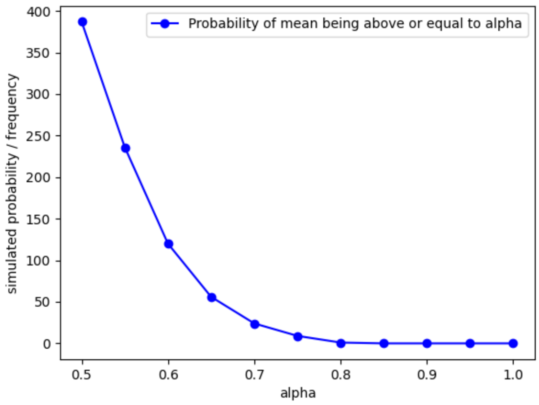

Figure 1: my captions

2

Since we take the mean of 20 Bernoulli random variables, the mean will only take on 21 different values, i.e. it will live on the space $\{0, \frac{1}{20}, \ldots, 1\}$. Due to having increments of size $\frac{1}{20}$ from adding another 'success' from one of the 20 realisations. Hence nothing new will be happening between the steps defined in the α vector. Of course if we were to increase n this would change, as the mean would be more fine-grained.

2 The Role of Independence (13 points)

An easy but instrictive example is to make the X_i's completely dependent of each other, that is $X_1 = X_n, n \geq 1$.
If we set $\mu = \mathbb{E}X_1 = 0.5$ then the mean will be either 0 or 1 for all $n \geq 1$ and the absolute difference to μ will hence be $\frac{1}{2}$ with probability 1.

Note that if we let $\mu \in (0,1)$ then the difference will only be greater than $\frac{1}{2}$ if the less likely outcome happens:

$$\mathbb{P}\left(|\mu - \frac{1}{n}\sum_{i=1}^{n} X_i| \geq \frac{1}{2}\right) = \min(\mu, 1-\mu)$$

3 Tightness of Markov's Inequality (Optional, question not for submission, 0 points)

4 The effect of scale (range) and normalization of random variables in Hoeffding's inequality (13 points)

5 Linear Regression (50 points)

```
# Example code
# Creating an example array
data = np.array([5, 2, 8, 1, 6])

# Calculating cumulative sum using cumsum
cumulative_sum = np.cumsum(data)
```