

# Assignment 2

Silvan Robert Adrian

December 4, 2018

## Contents

<b>1</b>	<b>Illustration of Hoeffding's Inequality</b>	<b>1</b>
1.1	Plotting Hoeffding . . . . .	1
1.2	Compare . . . . .	1
1.3	Exact probability . . . . .	1
<b>2</b>	<b>The effect of scale (range) and normalisation of random variables in Hoeffding's inequality</b>	<b>2</b>
<b>3</b>	<b>A Very Practical Question</b>	<b>3</b>
3.1	markov . . . . .	3
3.2	Chebyshev . . . . .	4
3.3	Hoeffding . . . . .	4
3.4	Non-vacuous . . . . .	4
<b>4</b>	<b>Probability in Practice ("The Airline Question")</b>	<b>4</b>
4.1	Question 1 . . . . .	4
4.2	Question 2 . . . . .	4
<b>5</b>	<b>Logistic Regression</b>	<b>5</b>
5.1	Cross entropy error measure . . . . .	5
5.2	Logistic regression loss gradient . . . . .	5
5.3	Logistic regression implementation . . . . .	5
5.4	Iris Flower Data . . . . .	5

## 1 Illustration of Hoeffding's Inequality

### 1.1 Plotting Hoeffding

I added the hoeffding bound to the following plot, which now includes all of the bounds including the empirical frequency.

To calculate the hoeffding bound I used  $e^{-2*n*(\epsilon^2)}$ , from which I ended up on the following plot of the bound (including the other plots from assignment 1):

### 1.2 Compare

Hoeffding bound overlaps at the beginning with markov but then fastly decreases, even faster then chebyshev which we compared in the first assignment already. We also can discover that hoeffding gets better with a higher threshold, even better then chebyshev.

### 1.3 Exact probability

From Assignment 1 we have the following probabilities: So before we got the probability for  $\alpha = 0.95$ :  $\Pr\left[\frac{1}{20} \sum_{i=1}^{20} X_i \geq 0.95\right] = \frac{21}{1024} \approx 0.0205078125$

and for  $\alpha = 1$ :  $\Pr\left[\frac{1}{20} \sum_{i=1}^{20} X_i \geq 1\right] = \frac{1}{1024} \approx 0.0009765625$

With hoeffding we just need to calculate it by using  $e^{-2*n*(\epsilon^2)}$ , for that I directly used my programmed function in python:

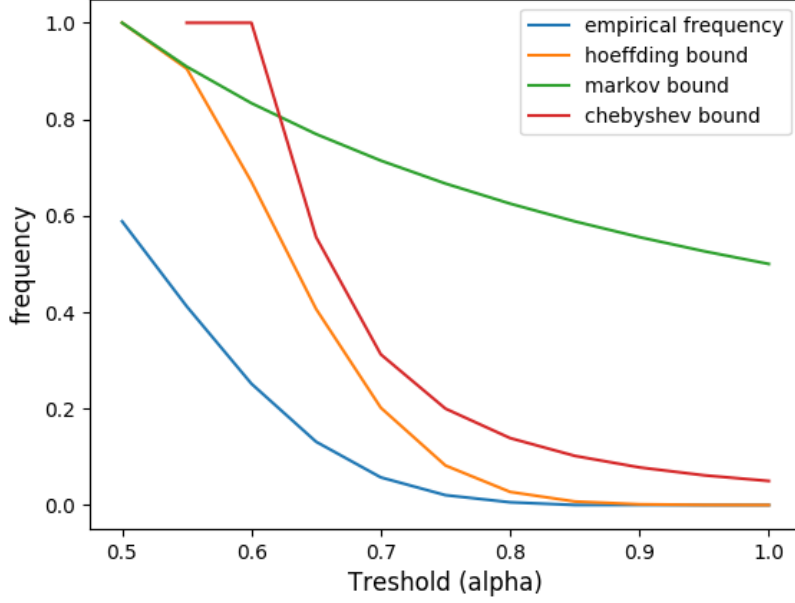


Figure 1: Plot with all bounds + the empirical frequency

then we get for  $\alpha = 0.95$ :  $hoeffding(0.5, 0.95, 20) \approx 0.0003035391380788673$   
and for  $\alpha = 1$ :  $hoeffding(0.5, 1, 20) \approx 4.5399929762484854e - 05$ .

So the probabilities lay quite near to each other. Which also gets clear from the plot above, so that hoeffding lands is very near to the exact probability. Which means for us the hoeffding estimation is very near the exact probability, so outputs quite good results in estimating the probabilities here.

## 2 The effect of scale (range) and normalisation of random variables in Hoeffding's inequality

Let all the assumptions of **corollary 2.5** be true for some random variables  $X_1, \dots, X_n$ . Set  $a_i = 0$  and  $b_i = 1$  for all  $i \in \{1, \dots, n\}$ . From **corollary 2.5** and our definition of  $a_i$  and  $b_i$ , we now have it that all assumptions of **theorem 2.3** are true for  $X_1, \dots, X_n$ . We can therefore use **theorem 2.3** to conclude that for all  $\epsilon > 0$ .

From **theorem 2.3** we have:

$$\mathbb{P} \left\{ \sum_{i=1}^n X_i - \mathbb{E} \left[ \sum_{i=1}^n X_i \right] \geq \epsilon \right\} \leq e^{-2\epsilon^2 / \sum_{i=1}^n (b_i - a_i)^2} \quad (1)$$

and

$$\mathbb{P} \left\{ \sum_{i=1}^n X_i - \mathbb{E} \left[ \sum_{i=1}^n X_i \right] \leq -\epsilon \right\} \leq e^{-2\epsilon^2 / \sum_{i=1}^n (b_i - a_i)^2} \quad (2)$$

The **corollary 2.5** states  $\mathbb{E}(X_i) = \mu$  for all  $i \in \{1, \dots, n\}$ , then we know that:

$$\mathbb{E} \left[ \sum_{i=1}^n X_i \right] = n\mu \quad (3)$$

Since  $b_i - a_i = 1$  for all  $i \in \{1, \dots, n\}$ , we also know that:

$$\sum_{i=1}^n (b_i - a_i)^2 = n \quad (4)$$

By using all the above learned we get for all  $\epsilon > 0$  following:

$$\mathbb{P} \left\{ \sum_{i=1}^n X_i - n\mu \geq \epsilon \right\} \leq e^{-2\frac{\epsilon^2}{n}} = e^{-2n(\frac{\epsilon}{n})^2} \quad (5)$$

and

$$\mathbb{P} \left\{ \sum_{i=1}^n X_i - n\mu \leq -\epsilon \right\} \leq e^{-2\frac{\epsilon^2}{n}} = e^{-2n(\frac{\epsilon}{n})^2} \quad (6)$$

By putting it then into the **corollary 2.5**, we see that for all  $\epsilon > 0$  follows that:

$$\mathbb{P} \left\{ \frac{1}{n} \sum_{i=1}^n X_i - \mu \geq \frac{\epsilon}{n} \right\} \leq e^{-2n(\frac{\epsilon}{n})^2} \quad (7)$$

and

$$\mathbb{P} \left\{ \frac{1}{n} \sum_{i=1}^n X_i - \mu \leq -\frac{\epsilon}{n} \right\} \leq e^{-2n(\frac{\epsilon}{n})^2} \quad (8)$$

Then it's clear that we can redefine the  $\epsilon$  as  $\tilde{\epsilon} = \frac{\epsilon}{n} > 0$ , so that we can conclude that for all  $\tilde{\epsilon} > 0$ :

$$\mathbb{P} \left\{ \frac{1}{n} \sum_{i=1}^n X_i - \mu \geq \tilde{\epsilon} \right\} \leq e^{-2n\tilde{\epsilon}^2} \quad (9)$$

and

$$\mathbb{P} \left\{ \frac{1}{n} \sum_{i=1}^n X_i - \mu \leq -\tilde{\epsilon} \right\} \leq e^{-2n\tilde{\epsilon}^2} \quad (10)$$

So we have proven that **corollary 2.5** follows from **theorem 2.3**.

### 3 A Very Practical Question

#### 3.1 markov

-

### 3.2 Chebyshev

-

### 3.3 Hoeffding

-

### 3.4 Non-vacuous

-

## 4 Probability in Practice (“The Airline Question”)

### 4.1 Question 1

We assume that the probability of each passenger showing up is 0.95 (since probability of not showing up is 0.05), and that the event of a passenger showing up is independent (which is the case according to the description). Which implies  $P[\text{passengershowingup}] = 0.95$ , from which we can calculate

$$\left(\frac{95}{100}\right)^{100} = 0.00592... \approx 0.006 \quad (11)$$

So for 6 out of 1000 flights there might show up too many passengers, so the plane won't have enough seats left for all passengers in those cases.

### 4.2 Question 2

We have given the following Events:

1. 95% of passengers showing up out of 10000, where each passenger shows up with the probability  $p$
2. 100 passengers show up each with the probability of  $p$  (everybody shows up)

From those 2 Events we get the probability that 9500 of 10000 persons getting on their plane and all 100 passengers arriving to one given airplane (overbooking). With this information we use the hoeffding bound for the probability of 9500 arriving to a plane. When this bound is found we then can multiply it by the probability that a plane gets overbooked ( $p^{100}$ ). Then I can solve for  $p$  and find the bound.

$$\mathbb{P}\left(\sum_{i=1}^{10000} X_i \geq 9500\right) \quad (12)$$

$$= \mathbb{P}\left(\sum_{i=1}^{10000} X_i - 10000 \cdot p \geq 9500 - 10000 \cdot p\right) \quad (13)$$

$$\leq e^{-2(10000(0.95-p))^2 / \sum_{i=1}^{10000} (b_i - a_i)^2} \quad (14)$$

$$= \leq e^{-2(10000(0.95-p))^2 / \sum_{i=1}^{10000} (0-1)^2} \quad (15)$$

$$= \leq e^{-2(10000(0.95-p))^2 / \sum_{i=1}^{10000} 1} \quad (16)$$

$$= \leq e^{-2(10000(0.95-p))^2 / 10000} \quad (17)$$

Now using the expression for the probability (bound) of all people arriving to the plane  $p^{100}$  the probability can be bound:

$$p^{100} \cdot e^{-2(10000(0.95-p))^2 / \sum_{i=1}^{10000} 1} = 0.006797 \quad (18)$$

When we solve the equation numerically we will find  $p = 0.9526$ .

Which would be the worst-case p, but since I didn't quite understand what kind of plot was expected by the description in the assignment, I don't have any plot.

## 5 Logistic Regression

### 5.1 Cross entropy error measure

-

### 5.2 Logistic regression loss gradient

-

### 5.3 Logistic regression implementation

-

### 5.4 Iris Flower Data

-