Understanding Empirical Average and Hoeffding’s Inequaltiy

HUANG, Hao Yu

A Report Submitted for the Course CMSC5728 Assignment#1 of
Master of Science
in
Computer Science

The Chinese University of Hong Kong
Oct 2024

Contents

Chapter 1

Derivations

1.1 Hoeffding's Inequaltiy

In this section, I will firstly introduce the corollary of Hoeffding's Inequaltiy. After the explanation of this theorem, I will focus on how to derive the confidence interval with a given confidence.

Theorem 1.1.1. *Assume that $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, ..., \mathcal{X}_n$ are independet random variables with mean $\mu_1, \mu_2, ..., \mu_n$, and $\mathcal{X}_i \in [a_i, b_i]$, then*

$$P(|\frac{1}{n}\sum_{i=1}^{n}(\mathcal{X}_i - \mu_i)| \geq \epsilon) \leq 2e^{-2(n\epsilon)^2/\sum_{i=1}^{n}(b_i-a_i)^2}, \forall \epsilon \geq 0$$

1.2 Confidence Interval

In this section, I will derive the confidence interval of the average value for some samples with a given confidence.

Firstly we assume the confidence.

$$confidence = 1 - \delta \tag{1.1}$$

And it also means that for half of the interval, the confidence of the remain interval is less equal than δ.

$$P(\frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i \geq \epsilon) \leq e^{-2(n\epsilon)^2/\sum_{i=1}^{n}(b_i-a_i)^2} = \delta/2 \tag{1.2}$$

Because the variables here are all independent and identically distributed(IID), so the average value of their distribution are all the same. And the differences between the upper bound and lower bound of the distribution of these samples are also the same. Hence we get 1.3, where μ is the average value of the distribution and ϵ is half width of the confidence interval centered by empirical average value. And we also assume that the upper bound and lower bound of the distribution of all variables is b and a respectively.

$$P(\frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_i - \mu \geq \epsilon) \leq e^{-2(n\epsilon)^2/n(b-a)^2} = \delta/2 \tag{1.3}$$

Now our goal is to find the ϵ with a given confidence $1-\delta$ and we can go home. We can derive that with the following equations 1.4,1.5,1.6.

$$e^{-2(n\epsilon)^2/n(b-a)^2} = \delta/2 \tag{1.4}$$

$$\epsilon = \sqrt{\frac{ln(\frac{\delta}{2}) \times n(b-a)^2}{-2n^2}} \tag{1.5}$$

$$\epsilon = \sqrt{\frac{ln(\frac{2}{\delta}) \times (b-a)^2}{2n}} \tag{1.6}$$

Now we get half of the width of the confidence interval centered by the empirical average of the samples. So finally we can get the confidence interval in equations 1.7 and 1.8.

$$\hat{\mu} - \epsilon \leq \mu \leq \hat{\mu} + \epsilon \tag{1.7}$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_i - \sqrt{\frac{ln(\frac{2}{\delta}) \times (b-a)^2}{2n}} \leq \mu \leq \frac{1}{n}\sum_{i=1}^{n}\mathcal{X}_i + \sqrt{\frac{ln(\frac{2}{\delta}) \times (b-a)^2}{2n}} \tag{1.8}$$

We are done.

Chapter 2

What Shrinkage is Like

2.1 Why Shrink

In this section, I will explain what is the shrinkage of the confidence interval and why that will happen. I will explain that from the perspective of intuition and math.

From the perspective of intuition, think of the confidence interval as a net that catches the true mean. With fewer samples, the net has to be wide to ensure it catches the true mean because there’ s more uncertainty. As you gather more samples, you become more confident about where the true mean is, so you can tighten the net. The interval shrinks because you have a better estimate of the true mean.

From the perspective of math, we can focus of the equation 1.8. With a fixed confidence, as we increase the number of samples, which is n, we can find that ϵ will decrease. And it means the width of the confidence value become smaller, which means the confidence interval shrinks.

2.2 Visualization

To make it more intuitive, I will give the visualization of three cases. And all three cased are under the assumption that the samples are all from the uniform distribution. And I set the confidence value is 95%.

For the visualization, the Y-axis will be the upper or lower bound of the average value, while the X-axis will be the number of samples.

2.2.1 Case1

I generate n random outcomes from a uniform distribution between $[0, 1]$. And to show the shrinkage, I will set the n with a range of $[1, 300]$. I set $b \leftarrow 1$, $a \leftarrow 0$.

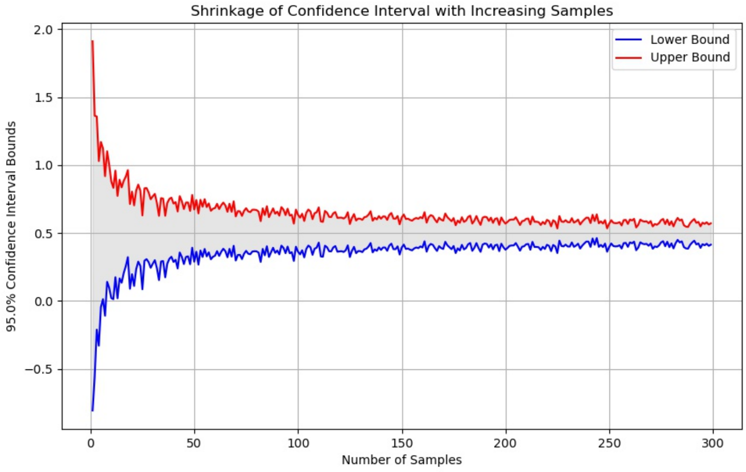

Figure 2.1: Case1

2.2.2 Case2

I generate n random outcomes from a uniform distribution between $[0, 2]$. And to show the shrinkage, I will set the n with a range of $[1, 300]$. I set $b \leftarrow 2$, $a \leftarrow 0$.

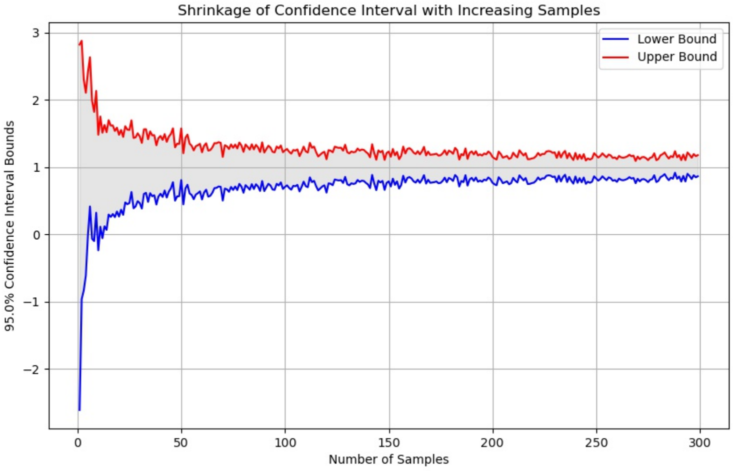

Figure 2.2: Case2

2.2.3 Case3

I generate n random outcomes from a uniform distribution between $[0,1]\cup[3,4]$. And to show the shrinkage, I will set the n with a range of $[1,300]$. I set $b \leftarrow 4$, $a \leftarrow 0$.

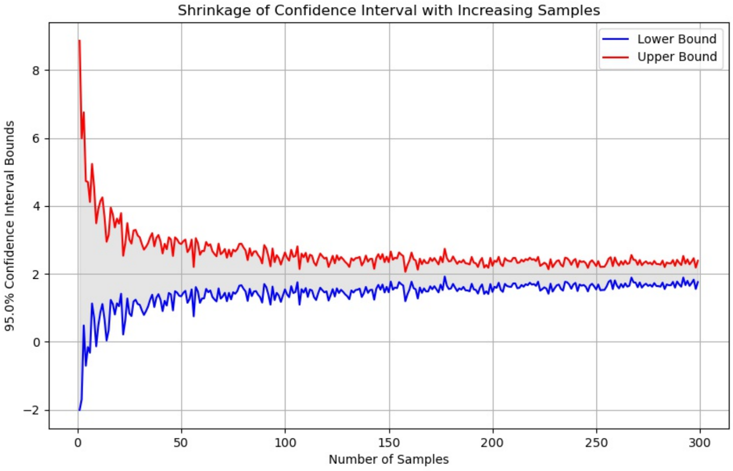

Figure 2.3: Case3

2.3 Analysis

As we can see from the last section, the confidence interval, which is the gray area, indeed shrinks as the number of samples increases.

Then I will give the derivations of the average value of the above threee distributions. Because they are all uniform distribution, so here firstly I give the derivation of the average value of uniform distribution. We assume that the interval is continuous, which is $[a, b]$. And the PDF is like equation 2.1.

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

For such a distribution, the expected average value is the weighted average of all possible values over the entire interval. Since a uniform distribution means that each value has an equal probability, we can calculate its mean using the following equation 2.3.

$$\mu = \int_a^b x \frac{1}{b-a} \, dx \tag{2.2}$$

And then we can get equation 2.5.

$$\mu = \frac{1}{b-a} \int_a^b x \, dx \tag{2.3}$$

$$\int_a^b x \, dx = \left[\frac{x^2}{2} \right]_a^b = \frac{b^2}{2} - \frac{a^2}{2} \tag{2.4}$$

$$\mu = \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{b+a}{2} \tag{2.5}$$

As for the case 3, the interval is not continuous, which is $[a, b] \cup [c, d]$. So we can derive that like the equation 2.6, 2.7, 2.8. Because if one interval is sampled more frequently, the corresponding distribution needs to be adjusted based on the weights. If the intervals are of different lengths or have different sampling probabilities, further weighting may be required.

$$\mu_1 = \frac{a+b}{2} \tag{2.6}$$

$$\mu_2 = \frac{c+d}{2} \tag{2.7}$$

$$\mu = \frac{(b-a) \times \mu_1 + (d-c) \times \mu_2}{(b-a)+(d-c)} \tag{2.8}$$

Then we can find that the value the interval shrinks to is just the average value of the distribution. And that also means our derivation is right.

2.4 Interesting Findings

However, I discover another phenomenon. While the sample size is very small, the confidence interval could even be larger than the interval of the distribution. For example, as figure 2.1 shows, when the number of samples is small the lower bound of the confidence could be smaller than zero. And the upper bound could also be larger than one. But I generate the number from a distribution of a uniform distribution between $[0, 1]$. That's very weird.

Actually, after some derivations, I found that it really could happen.

Let's say if the half of the width of the confidence interval is larger than the half of the width of the distribution interval, the phenomenon above will happen. So the condition is equation 2.9.

$$\frac{(b-a)}{2} \leq \epsilon = \sqrt{\frac{(b-a)^2 ln(\frac{2}{\delta})}{2n}} \tag{2.9}$$

With this equation 2.9, we can find what is the threshold of the sample size to avoid this phenomenon in equation 2.10.

$$n \leq 2(b-a)^2 ln(\frac{2}{\delta}) \tag{2.10}$$

And we can set the number in case1 to the equation 2.10. We will find the sample size should larger than $2 * (1 - 0)^2 ln(2/0.05)$, which is about 6, and the phenomenon will not happen. Because this phenomenon shows that if that happen, our samples is useless. We can not derive anything from the sampled data.

So the derivations above explain that when the number of sample numbers is lower than a threshold, our samples will be useless for estimating the distribution with Hoeffding's Inequality.