Lecture 21: The Chernoff-Hoeffding Bound

Anup Rao

May 20, 2019

Suppose Reuters wants to conduct a poll to understand how many people would vote for Joe Biden in the coming election. Say b fraction of all US citizens currently support Biden, and Reuters has the ability to call uniformly random citizens and ask them whether they would vote for Biden or not. If they call n people, and X say they will vote for Biden, a natural estimate is $b \approx X/n$. But how accurate is this estimate? How many people should Reuters call to be sure that this is a good estimate?

We have already seen a couple of ideas that are relevant to this question. Let $\theta = 0.05, \gamma = 0.01$ be parameters. To make the above question more concrete—our goal is to compute how large n needs to be so that

$$p(|X/n - b| \geq \theta) \leq \gamma.$$

If n satisfies the above equation, then we can assert that X/n is within θ of b, except with probability γ.

The interval $[X/n - \theta, X/n + \theta]$ is often called the *confidence interval*.

Using Variance and Chebyshev's Inequality

Chebyshev's inequality certainly gives us one way to solve this problem. We have that $X = X_1 + X_2 + \cdots + X_n$ is a binomial random variable, where each X_i is 1 with probability b. As we have seen before, $\text{Var}[X] = (b - b^2)n$, and then Chebyshev says:

$$\begin{aligned} p(|X/n - b| \geq \theta) &= p(|X - bn| \geq \theta n) \\ &\leq \frac{\text{Var}[X]}{(\theta n)^2} \\ &= \frac{b - b^2}{\theta^2 n} \leq \qquad\qquad = \frac{1}{\theta^2 n}. \end{aligned}$$

So, if we set $n \geq 1/(\gamma\theta^2)$, the probability is at most γ. For our choice of θ, γ, this gives that n needs to be at least $40,000$. That is a lot of phone calls.

Using the Central Limit Theorem

The central limit theorem says that if we take a large number of independent samples of a particular distribution, then the average of these samples corresponds to the corresponding normal. This gives us lots of information about the distribution about the average of independent random variables. In particular, one should expect that the probability that the average deviates from its expectation is exponentially small, because the normal distribution satisfies this fact:

This can be proved by integrating the pdf of the normal.

Fact 1 (Tail Bound for the Normal distribution). *Suppose X is a normal with mean μ and standard deviation σ. Then $p(|X-\mu| \geq k\sigma) \leq 2e^{-k^2/2}$.*

Using this estimate, we obtain:

Here $\mu = bn, \sigma = \sqrt{(b-b^2)n}$.

$$\begin{aligned} p(|X/n - b| \geq \theta) &= p\left(|X - bn| \geq \theta\sqrt{n/(b-b^2)} \cdot \sqrt{(b-b^2)n}\right) \\ &\lessapprox 2e^{-\frac{\theta^2 n}{2(b-b^2)}} \\ &\leq 2e^{-\frac{\theta^2 n}{2}}. \end{aligned}$$

since $b \leq 1$.

This is exponentially smaller than old bound, so n does not to be as large. Setting $n = 2 \cdot \ln(2/\gamma)/\theta^2$ is enough. For our choices of θ, γ, this gives $n \geq 4239$. This is much more reasonable, but there is one tiny problem—we do not know that it actually works, because the equations above have a $\lessapprox$ in them. The central limit theorem says that the distribution of X converges to the normal distribution for n that is *large enough*. But how large is large enough? Is $n = 4239$ large enough? The central limit theorem does not say anything about this.

In order to address this issue we use the Chernoff-Hoeffding bound.

Chernoff-Hoeffding

Suppose $X_1, \ldots, X_n$ are independent random variables taking values in between 0 and 1, and let $X = X_1 + X_2 + \ldots + X_n$ be their sum, and $\mathbb{E}[X] = \mu$. There are many forms of the Chernoff bounds, but here we focus on this one:

There are several other kinds of bounds like *Hoeffding* bounds and *Azuma's inequality* that are closely related to Chernoff bounds.

Theorem 2. *Suppose $0 < \delta$, then*

$$p(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2\mu}{2+\delta}},$$

and

$$p(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta^2\mu}{2}}.$$

You can combine both inequalities into one if you write it like this:

If $X_1, \ldots, X_n$ do not lie in between 0 and 1, you can always scale them so that they do, and then apply the bound. Careful though—scaling the random variables changes the value of μ as well!

Theorem 3. *Suppose $0 < \delta$, then*

$$p(|X-\mu| > \delta\mu) \leq 2e^{-\frac{\delta^2\mu}{2+\delta}}.$$

Now, let us apply this theorem to the polling problem. Setting $\delta = \theta/b$,

$$\begin{aligned} p(|X/n - b| \geq \theta) &= p(|X - bn| \geq \theta n) \\ &= p(|X - bn| \geq \delta bn) \\ &\leq 2e^{-\frac{\delta^2 bn}{2+\delta}} = 2e^{-\frac{(\theta/b)^2 bn}{2+\theta/b}} = 2e^{-\frac{\theta^2 n}{2b+\theta}} \leq 2e^{-\frac{\theta^2 n}{2+\theta}} \end{aligned}$$

So, if we want the probability of our estimate being off by θ to be at most γ, then we want

$$\begin{aligned} &2e^{-\frac{\theta^2}{2+\theta}\cdot n} \leq \gamma \\ \Rightarrow &e^{\frac{\theta^2}{2+\theta}\cdot n} \geq 2/\gamma \\ \Rightarrow &\frac{\theta^2}{2+\theta} \cdot n \geq \ln(2/\gamma) \\ \Rightarrow &n \geq \frac{2+\theta}{\theta^2} \cdot \ln(2/\gamma). \end{aligned}$$

For our choices of θ, γ, this gives $n \geq 4345$, nearly as good as the Central Limit Theorem!