SLLN example problem

Theorem 0.1 (SLLN). Suppose $X_1, X_2, \ldots$ are i.i.d. random variables with $\mathbb{E}[|X_i|] < \infty$, and define $S_n := \frac{1}{n} \sum_{i=1}^{n} X_i$ for all n. Then

$$S_n \to \mathbb{E}[X_1] \text{ almost surely.}$$

Here is an example problem for using the SLLN:

Problem 0.1. Let $X_1, X_2, \ldots$ be i.i.d. nonnegative random variables with finite mean $\mathbb{E}[X_1] = \lambda$. Fixing $\varepsilon > 0$, let C_m be the event that

$$\frac{1}{t} \sum_{i=1}^{t} X_i - \lambda \leq \varepsilon \text{ for all } t \geq m$$

Prove that there exists some m^* for which $\mathbb{P}[C_{m^*}] > 1/2$.

By the SLLN, we know that if we define the event A as

$$A := \omega : \lim_{t \to 0} \frac{1}{t} \sum_{i=1}^{t} X_i(\omega) - \lambda = 0$$

then $\mathbb{P}[A] = 1$. Let us also define the event B as

$$B := \bigcup_{m=1}^{\infty} C_m$$

Note that for any $\omega \in A$, there exists some integer $m(\omega)$ such that $C_{m(\omega)}$ happens (by definition). Therefore, for any $\omega \in A$, we know that $\omega \in B$ – and therefore $A \subseteq B$. But this implies $\mathbb{P}[\cup_m C_m] = \mathbb{P}[B] = 1$ (since $\mathbb{P}[A] = 1$).

But C_m are a nondecreasing sequence of events and therefore by continuity of probability $\lim_{m \to \infty} \mathbb{P}[C_m] = 1$. But that immediately means that for sufficiently large m^*, we have $\mathbb{P}[C_{m^*}] > 1/2$.

Convergence of empirical estimates, done three ways

Problem 0.2. Let $X_1, X_2, \ldots$ be i.i.d. $\sim \text{Bern}(p)$, and define $S_n := \frac{1}{n} \sum_{i=1}^n X_i$. We know by SLLN that $\lim_{n\to\infty} S_n \to p$ almost surely.

We now ask: given accuracy ε and confidence $1-\delta$, how many samples of X_i do we need to estimate p to that accuracy and confidence? Formally, given $\varepsilon, \delta > 0$, we want to know the smallest n such that

$$\mathbb{P}\left(|S_n - p| \leq \varepsilon\right) \geq 1 - \delta$$

This can be done three ways:

- Using Hoeffding to first show that

$$\mathbb{P}\left(|S_n - p| > \varepsilon\right) \leq 2e^{-\frac{n\varepsilon^2}{3}}$$

- Using Chebyshev's Inequality to first show that

$$\mathbb{P}\left(|S_n - p| > \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}$$

- Using the Central Limit Theorem, show that

$$\mathbb{P}\left(|S_n - p| > \varepsilon\right) \approx 2 - 2\Phi(2\varepsilon\sqrt{n})$$

  where Φ is the CDF of $\mathcal{N}(0,1)$ (to make this rigorous, use the Berry-Esseen Theorem).

We need the following theorems of course:

Theorem 0.2 (Hoeffding's Theorem). If X_i are i.i.d. Bernoulli random variables, and $X := \sum_{i=1}^n X_i$,

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq \alpha\mathbb{E}[X]\right) \leq 2e^{-\frac{\alpha^2}{3}\mathbb{E}[X]}$$

Theorem 0.3 (Chebyshev's Inequality). If X is a r.v. with finite variance, then

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq \alpha\right) \leq \frac{\text{Var}(X)}{\alpha^2}$$

Theorem 0.4 (Central Limit Theorem). Let $X_1, X_2, \ldots$ be i.i.d. with finite mean $\mathbb{E}[X_1] = \mu$ and finite variance $\text{Var}(X_1) = \sigma^2$. Then, defining $S_n := \frac{1}{n} \sum_{i=1}^n X_i$,

$$\frac{nS_n - n\mu}{\sqrt{n}\sigma} \to \mathcal{N}(0,1) \text{ (in distribution)}$$

By Hoeffding

We multiply everything by n since

$$\mathbb{P}\ |S_n - p| > \varepsilon\ = \mathbb{P}\ |nS_n - np| > n\varepsilon$$

nS_n is just the sum of the X_i's, and np their expected value. So we just plug into the Hoeffding theorem (with $\alpha = \varepsilon$) to get the result given. Now we need n large enough that

$$2e^{-\frac{n\varepsilon^2}{3}} \leq \delta$$

A little algebra reveals that this is equivalent to $n \geq \frac{3\log(2/\delta)}{\varepsilon^2}$

By Chebyshev

Again we multiply everything by n. Because the X_i are independent,

$$\mathrm{Var}(nS_n) = n\mathrm{Var}(X_i) = np(1-p) \leq \frac{n}{4}$$

Plugging in $\alpha = n\varepsilon$ and the variance above, we get

$$\mathbb{P}\ |S_n - p| > \varepsilon\ \leq \frac{1}{4n\varepsilon^2}$$

exactly as we wanted it. Again, we want n large enough that this is $\leq \delta$. Algebra gives: $n \geq \frac{1}{4\delta\varepsilon^2}$

By Central Limit Theorem

Once again, multiply everything by n (and move the denominator over) using $\mu = p$, and define $Y_n := \frac{nS_n - n\mu}{\sqrt{n}\sigma}$. Then we get

$$|Y_n| > \frac{\varepsilon\sqrt{n}}{\sigma} \iff \frac{nS_n - n\mu}{\sqrt{n}\sigma} > \frac{\varepsilon\sqrt{n}}{\sigma} \iff S_n - p > \varepsilon$$

But by the CLT, Y_n is approximately distributed according to $\mathcal{N}(0,1)$. Therefore, by symmetry,

$$\mathbb{P}\ |S_n - p| > \varepsilon\ = \mathbb{P}\ |Y_n| > \frac{\varepsilon\sqrt{n}}{\sigma}\ \approx 2\ 1 - \Phi\ \frac{\varepsilon\sqrt{n}}{\sigma}\ = 2 - 2\Phi\ \frac{\varepsilon\sqrt{n}}{\sigma}\ \leq 2 - 2\Phi(2\varepsilon\sqrt{n})$$

The last step happens because $\sigma^2 = p(1-p) \leq 1/4$, which implies $\sigma \leq 1/2$. Note that this is highly non-rigorous, so when in doubt double-check with Berry-Esseen.

Finally, let's see what n is required for a given ϵ, δ. We really want to upper-bound the probability of failure by δ, meaning:

$$2 - 2\Phi(2\varepsilon\sqrt{n}) = \delta \iff \Phi(2\varepsilon\sqrt{n}) = \frac{2-\delta}{2} \iff n = \ \frac{\Phi^{-1}\ \frac{2-\delta}{2}}{2\varepsilon}^{\ 2}$$

Conclusion

Note that the three results have equivalent dependencies on $\frac{1}{\varepsilon}$ (i.e. $\frac{1}{\varepsilon^2}$), but Hoeffding has much better dependence on $\frac{1}{\delta}$ than Chebyshev does; however, Chebyshev has a better constant and can therefore be better when confidence doesn't need to be super large. Meanwhile, the CLT-derived bound is by far the strongest in general but you have to be careful about when you can use it (again, for full rigor, apply Berry-Esseen). For instance, if $\delta = 0.05$ (95% confidence wanted) and $\varepsilon = 0.01$, then:

- Chebyshev proves that 50,000 trials are sufficient.
- Hoeffding proves that 110,667 trials are sufficient.
- CLT proves that (approximately) 9,604 trials are sufficient.

However, if we require much higher confidence, i.e. $\delta = 0.001$, then

- Chebyshev proves that 2,500,000 trials are sufficient.
- Hoeffding proves that 228,028 trials are sufficient.
- CLT proves that (approximately) 27,069 trials are sufficient.

The Chernoff-Union One-Two Punch Combo

Problem 0.3. Let $Z_1, \ldots, Z_n$ be uniformly distributed (i.i.d.) in $[0,1]^2$, and let $L(Z_1, \ldots, Z_n)$ be the length of the shortest continuous path which visits all n points. Prove that with high probability, $L_n \propto n^{1/2}$, i.e. that there are constants $0 < b < B$ and a polynomial (or faster-growing function) $q(n)$ such that

$$b\, n^{1/2} \leq L(Z_1, \ldots, Z_n) \leq B\, n^{1/2}$$

with probability at least $1 - 1/q(n)$ (for all sufficiently large n).

Fun fact: The proof we'll demonstrate also suffices to show that if we are in d-dimensional space, $L \propto n^{1-1/d}$; in fact, it can be generalized to show strong lower bounds for L when the path must satisfy differential constraints as well! (This was part of my master's thesis.)

The upper-bound, in about two seconds

The upper bound is easy and not really the focus here. We can just split up the square into n cells of size $n^{-1/2} \times n^{-1/2}$ and then travel from cell to cell getting all the points before moving on. This takes time proportional to $n^{1/2}$ because each arc has length at most $\sqrt{2}n^{-1/2}$ and there are at most $2n$ of them (point to point and cell to cell). Note that this *always* works.

The lower-bound game plan

For the lower bound, we ask a related question: what is the maximum number of points we can collect with a path of length 1? Clearly if the answer is at most proportional to $n^{1/2}$ we are done because it would then take at least (proportional to) $n^{1/2}$ such paths to get all n points.

We'll begin with the following facts. The proof would work with any constant, but 7 is the smallest constant which works and this makes the constant bounds tighter. The second bound is certainly not the tightest possible, but it doesn't matter (it doesn't even really make the constant bounds worse)!

Fact 0.1. For any ε, a radius-2ε circle can be covered by 7 radius-ε circles.

Fact 0.2. The unit square can be covered by ε^{-2} circles of radius ε.

Now we are going to follow this game plan:

- Discretize the problem by representing paths by sequences of (appropriately sized) circles.
- Compute the number of such sequences of circles (horrifyingly large!)
- Compute the probability that an arbitrary fixed sequence contains "too many" points (very small thanks to Chernoff!)
- Apply the Union Bound to get the result.

Discretization

Let $B_\ell(z)$ denote the ball of radius ℓ centered at z. We also fix a 'canonical' configuration of seven radius-ℓ circles for any radius-2ℓ circle.

Suppose we let $\epsilon = n^{-1/2}$, and cover the space with $\epsilon^{-2} = n$ circles of radius ε. Now we have a path of length ≤ 1 - call it a function $\phi : [0,1] \to [0,1]^2$ which satisfies the *Lipschitz condition* (i.e. $\|\phi(x_1) - \phi(x_2)\|_2 \leq |x_1 - x_2|$) – this is effectively a "speed limit" on the function.

Let us check in on ϕ every ε – i.e. we look at $\phi(t\,\varepsilon)$ for all $t = 1, 2, \ldots, n^{1/2}$. Specifically, we do the following:

1. Let ψ_0 be the center of the circle $\phi(0)$ falls into (if there is more than one, pick arbitrarily).

2. We build a sequence of points $\boldsymbol{\psi} := \psi_0, \psi_1, \ldots, \psi_{n^{1/2}-1}$ as follows:

   (a) For every $t = 0, 1, \ldots, n^{1/2} - 1$, if $\phi(t\,\varepsilon) \in B_\varepsilon(\psi_t)$, then $\phi((t+1)\,\varepsilon) \in B_{2\varepsilon}(\psi_t)$ (can't make it out in time because of the Lipschitz condition).

   (b) Therefore, by Fact 1 from the previous page, it must be in one (at least - if more than one, pick arbitrarily) of the 7 radius-ε circles which cover $B_{2\varepsilon}(\psi_t)$ – so let ψ_{t+1} be the center of that circle.

   Note that this always preserves the condition that $\phi(t\,\varepsilon) \in B_\varepsilon(\psi_t)$.

Note also that between $\phi(t\,\varepsilon)$ and $\phi((t+1)\,\varepsilon)$, the path can never leave $B_{2\varepsilon}(\psi_t)$ (by the Lipschitz condition) – and therefore the path is entirely contained in

$$S_{\boldsymbol{\psi}} := \bigcup_{t=0}^{n^{1/2}-2} B_{2\varepsilon}(\psi_t)$$

So how many different sequences of this type are possible? Well, denote the set of all such sequences of Ψ:

- There are n choices for the ψ_0 (just the n circles covering $[0,1]^2$).

- For each of $n^{1/2}$ steps, there are (at most since we might be on the boundary) 7 choices for the next circle.

Therefore $|\Psi| = n \cdot 7^{n^{1/2}}$, which we'll just round up to $e^{2n^{1/2}}$ because frankly we're not trying to make this super efficient anyway and $e^2 > 7$ is convenient.

Now what we'll do is see the maximum number a *sequence of circles* can cover; since every length-1 path is contained in such a sequence this upper-bounds the number of points any length-1 path can collect.

A fixed circle-sequence

Now let's fix an arbitrary sequence $\boldsymbol{\psi} = \psi_0, \ldots, \psi_{n^{1/2}-1}$ *before* the random targets are assigned and ask: how many random targets does this sequence cover (with circles of radius 2ε)?

Well, the sequence consists of $n^{1/2}$ circles of area $\pi(2\varepsilon)^2 = 4\pi n^{-1}$. Thus, the area is at most $4\pi n^{-1/2}$ (again, we can make this more efficient by noting that the circles must overlap and yada yada, but that's a lot of work and doesn't really change the point). Let's define the random variables

$$X_i := \begin{cases} 1 & \text{if } Z_i \in S_{\boldsymbol{\psi}} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad X_{\boldsymbol{\psi}} := \sum_{i=1}^{n} X_i$$

Since the points are dropped in uniformly, we get

$$\mathbb{E}[X_{\boldsymbol{\psi}}] := \mathbb{E}[\#\{i : Z_i \text{ in } S_{\boldsymbol{\psi}}\}] = \sum_{i=1}^{n} X_i \leq 4\pi n^{-1/2} \cdot n = 4\pi n^{1/2}$$

Note that equality is "worst case" here, so if we treat the area as equal to this, we'll get an upper-bound, as we want. Since the points are independent, we get to apply the *Chernoff Bound*. Specifically, we use this version:

Theorem 0.5 (Hoeffding Upper Bound). If X_i are i.i.d. Bernoulli random variables, and $X := \sum_{i=1}^{n} X_i$,

$$\mathbb{P}\left(X \geq (1+\delta)\mathbb{E}[X]\right) \leq e^{-\frac{\delta^2}{3}\mathbb{E}[X]}$$

Plugging in our $\mathbb{E}[X]$, we get that

$$\mathbb{P}\left(X_{\boldsymbol{\psi}} \geq (1+\delta)4\pi n^{1/2}\right) \leq e^{-\frac{\delta^2}{3}4\pi n^{1/2}} = e^{-\frac{4\pi}{3}\delta^2 n^{1/2}}$$

We will determine what δ we should use later - for now, let's just keep it as a variable.

Union-bounding, and choosing δ

Okay, now we simply combine the two results from above with the Union Bound, define an appropriate δ, and finish. We are interested in bounding

$$\mathbb{P}[\max_{\boldsymbol{\psi} \in \Psi} X_{\boldsymbol{\psi}} \geq (1+\delta)4\pi n^{1/2}]$$

We have $e^{2n^{1/2}}$ sequences in Ψ; each has (at most) a probability of $e^{-\frac{4\pi}{3}\delta^2 n^{1/2}}$ to break the bound. Therefore, the above is just

$$\mathbb{P}[\max_{\boldsymbol{\psi} \in \Psi} X_{\boldsymbol{\psi}} \geq (1+\delta)4\pi n^{1/2}] \leq e^{2n^{1/2}} e^{-\frac{4\pi}{3}\delta^2 n^{1/2}}$$

Let's pick $\delta = 0.7$, giving us

$$\mathbb{P}[\max_{\boldsymbol{\psi} \in \Psi} X_{\boldsymbol{\psi}} \geq 1.7 \cdot 4\pi \cdot n^{1/2}] \leq e^{2n^{1/2}} e^{-\frac{4\pi}{3}(0.7)^2 n^{1/2}} < 2e^{-0.052 \cdot n^{1/2}}$$

Finishing the argument

Since $\max_{\psi \in \Psi} X_{\psi}$ is an upper bound for how many points a length-1 path can collect, we get the result that

Proposition 0.1. With probability at least $1 - e^{-0.052 \cdot n^{1/2}}$, there is no length-1 path which collects more than $1.7 \cdot 4\pi \cdot n^{1/2} < 22\, n^{1/2}$ of the n random points.

This now finishes what we wanted to show about the TSP, because if it takes length 1 to get $22\, n^{1/2}$ points, it will take at least a path of length $\frac{1}{22} n^{1/2}$ to get all n of them giving our final theorem (combined with the upper bound at the top):

Theorem 0.6. Let $Z_1, \ldots, Z_n$ be uniformly distributed (i.i.d.) in $[0,1]^2$, and let $L(Z_1, \ldots, Z_n)$ be the length of the shortest continuous path which visits all n points. Then

$$\frac{1}{22}\, n^{1/2} \leq L(Z_1, \ldots, Z_n) \leq 2\sqrt{2}\, n^{1/2}$$

with probability at least $1 - e^{-0.052 \cdot n^{1/2}}$ (for all sufficiently large n).

Note that the constant factors here are rather horrendous. They can be tightened quite a bit on both ends; but fundamentally this technique will suffer from this problem because of how generous we're prepared to be to allow the paths to get *all* the points in their associated sets. But it does show rate-of-growth quite nicely and can be generalized to a much wider class of problems in which the paths have to satisfy differential constraints.