CSE 840: Computational Foundations of Artificial Intelligence	November 27, 2023
Lecture 22	
Instructor: Vishnu Boddeti	*Scribe: Matthew McDerment*

1 Limit Theorems: LLN and CLT

1.1 Strong Law of Large Numbers

$X_n : (\Omega, A, P) \to \mathbb{R}$ i.i.d (independent and identically distributed). Assume the mean $\mu := E(X_n) < \infty$, and $Var(X_n) =: r^2 < \infty$. Then: $\lim_{x\to\infty} \frac{1}{n}\sum_{i=1}^{n} X = \mu a.s.$ and in L^2.

Examples: Train error, test error. Converge to the true error. In statistics, compare if means of two distributions are the same.

1.2 Weak Law of Large Numbers:

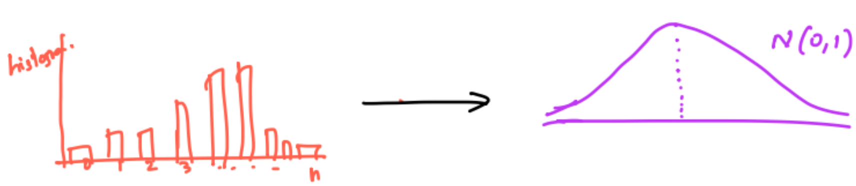
Converge in probability
Remarks: Many versions of this theorem exist (slightly relaxing i.i.d)

- "Strong law" ⇆ Convergence $a.s.$
- "Weak law" ⇆ Convergence in probability.
- There are cases where this fails, e.g. heavy failed distributions.
- If there is a selection bias in my samples (typical in human economic/rational behavior) the LLN does not mitigate the bias.

1.3 Central Limit Theorem

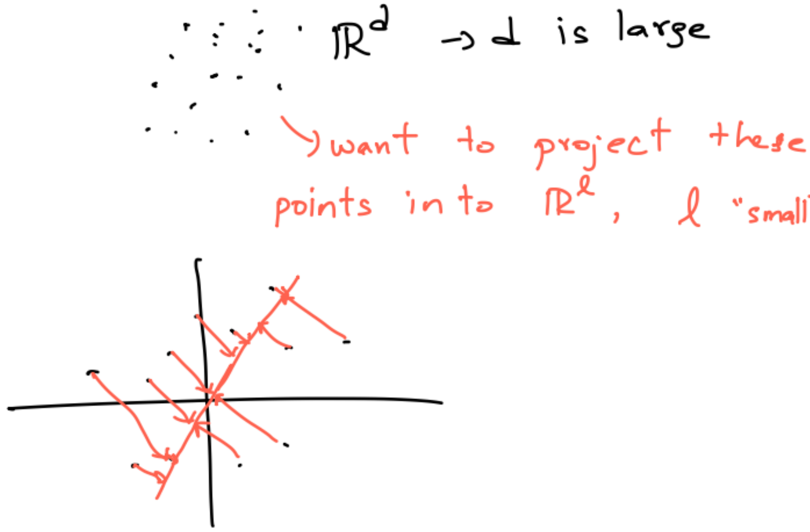
$(X_i)_{i\epsilon\varnothing}$ i.i.d random variables with mean μ and variance $r^2 < \infty$. Consider the RV $S_n : \sum_{i=n}^{n} X_i$. WE normalize it to $Y_n := \frac{S_n - n*\mu}{\sqrt{n}r}$ (Which has mean 0 and std. deviation 1). Then $Y_n \to Y$ in distribution where $Y\ N(0,1)$

Illustration: X_i coin, head = 1, tail = 0 $S_n = \sum X_i \epsilon [0, n]$



2 Concentration Inequalities

Motivation: Random projections



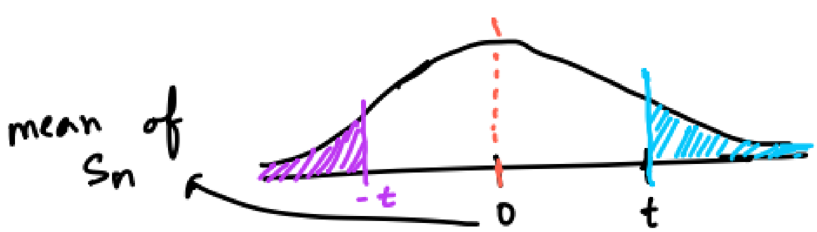
2.1 Theorem of Johnson-Lindenstrauss:

Can guarantee (for certain parameters ε, R)
$(1-\varepsilon)||x_i - x_j||_{\mathbb{R}}d \leq ||\pi(x_i) - \pi(x_j)||_{\mathbb{R}}l \leq (1+\varepsilon)||x_i - x_j||\mathbb{R}v$ Constructon/Proof steps:

- Assume you know $||x_i - x_j||_{\mathbb{R}}d = 1$. Compute $E(||\pi(x_i) - \pi(x_j)||_{\mathbb{R}}l)$
- $P(|(||\pi(x_i) - \pi(x_j)|| - E(...))| > t)$?

3 Hoeffding Inequality

Theorem 1 ***Hoeffding:*** *$x_1...x_n : (\omega, A, P) \to (\mathbb{R}, B)$ RVs, independent, assume that $X_i \epsilon [a_i, b_i] a.s.$ for $i = 1, 2, ...n$. Let $S_n := \sum_{i=1}^{n}(x_i - E(x_i))$. Then for any $t > 0$, $P(S_n \geq t) \leq exp(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)})$*



3.1 Application of Hoeffding: SLLN

Prop: $(X_i)_{i\epsilon\varnothing}$ i.i.d. RV, $a \leq x_i b$, let x have the same distribution as the x_i then $\frac{1}{n}\sum_{i=1}^{n} x_i \to Exa.s.$

Proof: Hoeffding $\to$

- $P(\frac{1}{n}\sum_{i=1}^{n} x_i - E(x) > t) \leq exp(\frac{-2(nt)^2}{\sum_{i=1}^{n}(b-a)^2}) = exp(\frac{-2nt^2}{(b-a)^2})$
- $P(\frac{1}{n}\sum x_i - E(x) < t) = P(\frac{1}{n}\sum(-x_i) - E(-x) > t) \leq exp(\frac{-2nt^2}{(b-a)^2})$

Combining the two, we get
$P(|\frac{1}{n}\sum x_i - E(x)| > t) \leq 2exp(-\frac{2nt^2}{(b-a)^2})$.
Now we want to apply Borel-Cantelli to get a.s. convergence: $\mathbb{Z} = \frac{1}{n}\sum_{i=1}^{n} x_i$
$\sum_{n=0}^{\infty} P(\mathbb{Z}_n - E(x) > t) \leq 2 * r \leq \infty$

- Substitute $r := exp(\frac{-2t^2}{(b-a)^2})\epsilon[0,1]$
- Observe: $exp(\frac{-2nt^2}{(b-a)^2}) = r^n$
- Sum: $2\sum_{n=0}^{\infty} r^n = 2 * \frac{1}{1-r} < \infty$

Now Borel-Cantelli gives almost sure convergence. □
Remark: Hoeffding is tight (cannot be improved without further assumptions). For fair coin tosses it is tight. But not tight if coin is biased $\to$ need other inequalities.

4 Bernstein Inequality

Theorem 2 ***Bernstein:*** *$x_1, ...x_n$, independent with 0 mean, $|x_i| < 1 a.s.$ Let $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} var(x_i)$. Then for all $t > 0$,*
$P(\frac{1}{n}\sum_{i=1}^{n} x_i > t) \leq exp(\frac{-nt^2}{2(\sigma^2 + \frac{t}{3})})$

5 Concentration Inequality For Funcs. With Bounded Difference

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ (or more generally, $f : x^n \to \mathbb{R}$ for some arbitrary space x).
We say that f has the bounded difference property if there xists constants $c_1, c_2...c_n$ such that $x_1...x_n\epsilon x|f(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n)\tilde{x}\epsilon x - f(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n)| \leq c_i$
Example: $f(x_1...x_n) = \sum_{i=1}^{n} x_i$ and $a \leq x_i \leq b$ Vi, then f satisfies with $c_i = b - a$.

Theorem 3 ***Mcdiarmid:*** *$x_1, ..., x_n$ independent RV; $x_i\epsilon x_i$, $f : x_1 * x_2, ...x_n \to \mathbb{R}$ function with bounded difference property. Then, for any $t > 0$,*
$P(f(x_1, x_2.., x_n) - E(f(x_1, x_2...x_n)) > t) \leq exp(\frac{-2t^2}{\sum_{i=1}^{n} C_i^2})$

Applications:

- Leave one out of error estimates
- Stability in ML
- Standard theoretical CS, randomized algos. (eg. traveling salesman problem)
- Largest eigenvalue of random symmetric matrices