Limit Theorems : LLN and CLT

Strong Law of Large Numbers

$X_n : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ i.i.d. (independent and identically distributed). Assume the mean $\mu := E(X_n) < \infty$, and $Var(X_n) =: \sigma^2 < \infty$.

Then : $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu$ 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.

Weak Law of Large Numbers : convergence in probability.

Remarks: · Many versions of this theorem exist (slightly relaxing i.i.d)

· "Strong law" $\Leftrightarrow$ convergence a.s.

· "weak law" $\Leftrightarrow$ convergence in probability.

⚠ • There are cases where this fails
e.g. heavy tailed distributions.

• If there is a selection bias in my samples
(typical in human economic/rational behavior)
then LLN does not mitigate the bias.

Central Limit Theorem

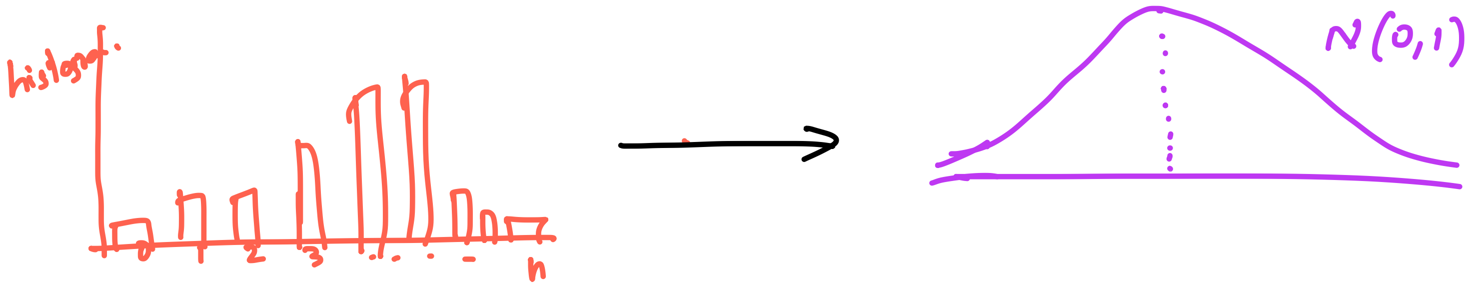
$(X_i)_{i \in \mathbb{N}}$ i.i.d. random variables with mean μ and variance $\sigma^2 < \infty$. Consider the RV $S_n := \sum_{i=1}^{n} X_i$. We normalize it to

$$Y_n := \frac{S_n - n \cdot \mu}{\sqrt{n}\,\sigma} \quad \left(\begin{array}{l}\text{which has mean } 0 \\ \text{and std. deviation } 1\end{array}\right)$$

Then $Y_n \to Y$ in distribution wher $Y \sim \mathcal{N}(0,1)$.

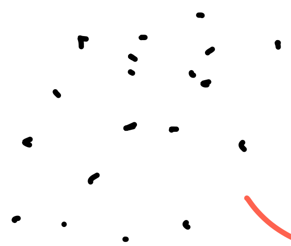
Illustration: X_i coin, head $= 1$, tail $= 0$

$S_n = \sum X_i \in [0, n]$



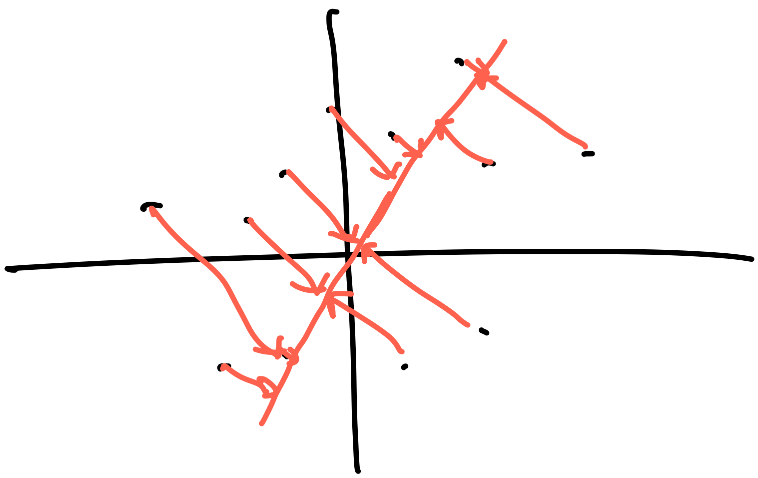
Concentration Inequalities

Motivation: random projections



$\mathbb{R}^d \to d$ is large

want to project these points in to $\mathbb{R}^\ell$, ℓ "small"



Theorem of Johnson–Lindenstrauss:

Can guarantee (for certain parameters ε, k)

$$(1-\epsilon)\,\|x_i - x_j\|_{\mathbb{R}^d} \leq \|\Pi(x_i) - \Pi(x_j)\|_{\mathbb{R}^\ell}$$

$$\leq (1+\epsilon)\,\|x_i - x_j\|_{\mathbb{R}^d}$$

Construction / proof steps:

(i) Assume you know $\|x_i - x_j\|_{\mathbb{R}^d} = 1$.

Compute $E\left(\|\pi(x_i) - \pi(x_j)\|_{\mathbb{R}^\ell} \right)$

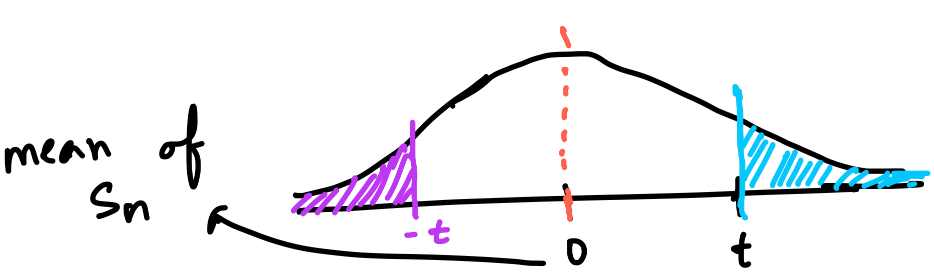
(ii) $P\left(\Big| \|\pi(x_i) - \pi(x_j)\| - E(\ldots\ldots) \Big| > t \right)$?

Hoeffding Inequality

Theorem (Hoeffding): $X_1 \ldots X_n : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ RVs, independent, assume that $X_i \in [a_i, b_i]$ a.s. for $i = 1, 2, \ldots n$. Let $S_n := \sum_{i=1}^{n} (X_i - E[X_i])$.

Then for any $t > 0$,

$$P(S_n \geq t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{n} (b_i - a_i)} \right)$$



Application of Hoeffding: SLLN

Prop: $(X_i)_{i\in\mathbb{N}}$ i.i.d. RV, $a \le X_i \le b$, let X have the same distribution as the X_i

Then $\frac{1}{n}\sum_{i=1}^{n} X_i \to E(X)$ a.s.

Proof: Hoeffding $\Rightarrow$

- $$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i - E(X) > t\right) \le \exp\left(\frac{-2(nt)^2}{\sum_{i=1}^{n}(b-a)^2}\right)$$
$$= \exp\left(\frac{-2nt^2}{(b-a)^2}\right)$$

- $$P\left(\frac{1}{n}\sum X_i - E(X) < -t\right) = P\left(\frac{1}{n}\sum(-X_i) - E(-X) > t\right)$$
$$\le \exp\left(\frac{-2nt^2}{(b-a)^2}\right)$$

Combining the two, we get

$$P\left(\left|\frac{1}{n}\sum X_i - E(X)\right| > t\right) \le 2\exp\left(-\frac{2nt^2}{(b-a)^2}\right).$$

Now we want to apply Borel-Cantelli to get a.s. convergence: $Z_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$\sum_{n=0}^{\infty} P(Z_n - E(X) > t) \leq 2 \cdot \underbrace{\exp\left(-\frac{2nt^2}{(b-a)^2}\right)}_{*} \overset{!}{<} \infty$$

$*$ substitute $\gamma := \exp\left(-\frac{2t^2}{(b-a)^2}\right) \in [0, 1)$

Observe: $\exp\left(\frac{-2nt^2}{(b-a)^2}\right) = \gamma^n$

sum: $2 \sum_{n=0}^{\infty} \gamma^n = 2 \cdot \frac{1}{1-\gamma} < \infty$

Now Borel-Cantelli gives almost sure convergence. ∎

<u>Remark</u>: Hoeffding is tight (cannot be improved without further assumptions).

For fair coin tosses it is tight.

<u>But</u>: not tight if coin is biased $\rightsquigarrow$ need other inequalities.

Berstein Inequality

Theorem (Bernstein): $X_1, \dots X_n$, independent with 0 mean, $|x_i| < 1$ a.s. Let $\sigma^2 := \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(x_i)$.

Then for all $t > 0$,

$$P\left(\frac{1}{n}\sum_{i=1}^{n} x_i > t\right) \leq \exp\left(\frac{-nt^2}{2(\sigma^2 + t/3)}\right)$$

Concentration inequality for funcs. with bounded difference

Consider a function $f: \mathbb{R}^n \to \mathbb{R}$ (or more generally, $f: \mathcal{X}^n \to \mathbb{R}$ for some arbitrary space $\mathcal{X}$).

We say that f has the bounded difference property if there exists constants $c_1, c_2 \dots c_n$ such that,

$$\sup_{\substack{x_1, \dots, x_n \in \mathcal{X} \\ \tilde{x}_i \in \mathcal{X}}} \Big| f(x_1, \dots, x_{i-1}, x_i, x_{i+1} \dots x_n) - f(x_1 \dots x_{i-1}, \tilde{x}_i, x_{i+1}, \dots x_n) \Big| \leq c_i$$

Example: $f(x_1 \ldots x_n) = \sum_{i=1}^{n} x_i$, and $a \leq x_i \leq b \;\forall i$, then f satisfies with $c_i = b-a$.

Theorem (McDiarmid): $X_1, \ldots, X_n$ independent RV; $X_i \in \mathcal{X}_i$, $f: \mathcal{X}_1 \times \mathcal{X}_2, \ldots \times \mathcal{X}_n \to \mathbb{R}$ function with bounded difference property. Then, for any $t > 0$,

$$P\Big(f(x_1, x_2 \ldots, x_n) - E\big(f(x_1, x_2 \ldots x_n)\big) > t \Big) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2} \right)$$

Applications:

- leave one out error estimates
- stability in ML
- standard theoretical CS, randomized algos. (e.g. traveling salesman problem)
- largest eigenvalue of random symmetric matrices.