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 \(\text{Var}(X_n) =: r^2 < \infty\). 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" \(\Rightarrow\) Convergence a.s.
- "Weak law" \(\Rightarrow\) 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

\((X_i)_{i \in \mathbb{N}}\) i.i.d random variables with mean \(\mu\) 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 \(Y \sim N(0,1)\)

**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 $\varepsilon, R$)

$$(1 - \varepsilon)||x_i - x_j||_d \leq ||\pi(x_i) - \pi(x_j)||_l \leq (1 + \varepsilon)||x_i - x_j||_d$$

Construction/Proof steps:

- Assume you know $||x_i - x_j||_d = 1$. Compute $E(||\pi(x_i) - \pi(x_j)||_l)$
- $P(||(||\pi(x_i) - \pi(x_j)|| - E(...)) > t)$

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 $t > 0$, $P(S_n \geq t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right)$
3.1 Application of Hoeffding: SLLN

**Prop:** \((X_i)_{i \in \mathbb{Z}}\) i.i.d. RV, \(a \leq x_i \leq b\), let \(x\) have the same distribution as the \(x_i\) then
\[
\frac{1}{n} \sum_{i=1}^{n} x_i \to \text{Exa.s.}
\]

**Proof:** Hoeffding \(ightarrow\)

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

Combining the two, we get
\[
P(\frac{1}{n} \sum_{i=1}^{n} x_i - E(x) > t) \leq 2 \exp\left(\frac{-2nt^2}{(b-a)^2}\right).
\]

Now we want to apply Borel-Cantelli to get a.s. convergence: \(Z = \frac{1}{n} \sum_{i=1}^{n} x_i\)
\[
\sum_{n=0}^{\infty} P(Z_n - E(x) > t) \leq 2 \ast r \leq \infty
\]

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

Now Borel-Cantelli gives almost sure convergence. \(\square\)

**Remark:** Hoeffding is tight (cannot be improved without further assumptions). For fair coin tosses it is tight. But not tight if coin is biased \(\rightarrow\) need other inequalities.

4 Bernstein Inequality

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

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 exists constants \(c_1, c_2, \ldots c_n\) such that
\[
x_1, \ldots, x_n \in [f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)] \leq c_i
\]

**Example:** \(f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i\) and \(a \leq x_i \leq b\) \(\forall i\), then \(f\) satisfies with \(c_i = b - a\).

**Theorem 3** **Mc diarmid:** \(x_1, \ldots x_n\) independent RV; \(x_i \in C\), \(f : x_1 \ast x_2, \ldots x_n \to \mathbb{R}\) function with bounded difference property. Then for any \(t > 0\),
\[
P(f(x_1, x_2, \ldots, x_n) - E(f(x_1, x_2, \ldots, x_n)) > t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{n} c_i^2}\right)
\]
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