Inequalities

- Many important models are too complex to analyze exactly.
- Help us prove important limiting theorems.

We will discuss Markov, Chebyshev & Hoeffding's bound.

**Markov's Inequality:**

For any non-negative RV $X$ and $t > 0$, we have

$$\Pr\{X \geq t^2\} \leq \frac{\mathbb{E}[X]}{t^2}$$

& tail probability (one-sided)

2. $X$

Average height of Stanford students is 165 cm, 5 feet 5 inches.

Fraction of students that are above 180 cm, 5 feet 11 inches.

$$P\{X \geq 180\} \leq \frac{165}{180}$$

$X = X \mathbb{I}_{\{X \geq t\}} + X \mathbb{I}_{\{X < t\}}$

$$\mathbb{E}[X] = \mathbb{E}[X \mathbb{I}_{\{X \geq t\}}] + \mathbb{E}[X \mathbb{I}_{\{X < t\}}]

\geq \int_{-\infty}^{t} x f_X(x) \, dx + \int_{t}^{\infty} x f_X(x) \, dx + \int_{-\infty}^{t} f_X(x) \, dx + \int_{t}^{\infty} f_X(x) \, dx$$

$$\geq t \int_{t}^{\infty} f_X(x) \, dx = t \Pr\{X \geq t\}$$
\[ \Pr\{X \geq t\} \leq \frac{E[X]}{t} \]

Chebyshev's inequality. 

\* better quadratic dependence
\* quantifies concentration
\* of X about mean, (controls both tails)

Let X be a r.v. with mean \( \mu \) and variance \( \sigma^2 \). Then for any \( t > 0 \), we have

\[ \Pr\{|X - \mu| \geq t\} = \Pr\left\{ \frac{(X - \mu)^2}{\sigma^2} \geq \frac{t^2}{\sigma^2} \right\} \leq \frac{E[(X - \mu)^2]}{t^2} = \frac{\sigma^2}{t^2}. \]

\[ \lim_{t \to \infty} \frac{1}{t^2} \int_0^t f(x) \, dx = 0. \]

**Limit Theorems:**

1) The Weak Law of Large Numbers (WLLN)

Let \( X_1, X_2, \ldots, X_n \) i.i.d. r.v.'s with mean \( \mu = E[X_1] = \cdots = E[X_n] \)

\[ S_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \overline{X}_n = \frac{1}{n} S_n. \quad \overline{X}_n \to \mu \text{ as } n \to \infty. \text{ WLLN} \]

\[ \forall \epsilon > 0 \quad \Pr\left\{ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| > \epsilon \right\} \to 0. \]

**Ex:** \( X_i \) independent flips of a coin with \( \Pr(X_i = 1) = 0.3 \)

**Ex:** Statistics / survey sampling.

**Ex:** Reliable communication.

Codebook = \( \mathcal{F} \) a collection of binary strings that are well separated from each other.

\[ 0.3 \quad 0.1 \]

\[ 0.1 \quad 0.3 \]

\[ 1 \quad 0 \]

\[ 0 \quad 1 \]

\[ 1 \quad 1 \]


\[ C = \{00000, 01111, 11100, 11011\} \]

use a codebook with codewords of length \( n \) for large \( n \) and make sure all codewords are separated from each other with a distance \( 2(0.1 + \varepsilon) n \) for some small \( \varepsilon \)

\[
\# \text{ of errors} \to 0.1
\]

\[ \frac{u}{n} \]

\[ \text{e.x.} \]

**Monte Carlo Integration**

\[
I(f) = \int_1^1 f(x) \, dx \to \text{numerical int. riemann sum}
\]

\[
\hat{I}(f) = \frac{1}{n} \sum_{i=1}^{n} f(X_i)
\]

\[
\hat{I}(f) \to \mathbb{E}[f(X_i)] = \int_0^1 f(x) \, dx
\]

\[
\frac{1}{n} \mathbb{E}\left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[X_i] = \mu
\]

\[
\mathbb{E}\left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[X_i] = \mu
\]

\[
\mathbb{E}\left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[X_i] = \mu
\]

\[
\text{Chebyshev's Inequality}
\]

\[
P\left( \left| \bar{X}_n - \mu \right| > \varepsilon \right) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} \]

\[
= \frac{\sigma^2}{n \varepsilon^2} \to 0 \quad \text{as} \quad n \to \infty
\]
\[ S_n = \frac{1}{n} \sum_{i=1}^{n} X_i \]

\[ \frac{S_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - n \mu}{\frac{\sigma}{\sqrt{n}}} \]

\[ \text{Var} \left[ \frac{S_n - n \mu}{\frac{\sigma}{\sqrt{n}}} \right] = \frac{1}{n} \sigma^2 \]

\[ W_n = \frac{\sqrt{n}}{\sigma} \left( \frac{S_n - n \mu}{\frac{\sigma}{\sqrt{n}}} \right) \]

\[ \text{Var} (W_n) = \frac{n}{n} \frac{1}{\sigma^2} \sigma^2 = 1 \]

\[ = \frac{S_n - n \mu}{\sqrt{n} \sigma} \xrightarrow{n \to \infty} \mathcal{N}(0, 1) \]

\[ E[W_n] = 0 \]

\[ \text{Var}(W_n) = 1 \]

Central Limit Theorem: Let \( X_1, X_2, \ldots, X_n \) iid with mean \( \mu \) and variance \( \sigma^2 \).

Let \( W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - n \mu) \) be their standardized sum.

Then \( \text{cdf of } W_n \xrightarrow{n \to \infty} \text{cdf of } \mathcal{N}(0, 1) \) for every \( \omega \); where \( W \sim \mathcal{N}(0, 1) \).

CLT: some justification for Gaussian models.

\[ \text{noise} = \text{sum of many independent small noise sources}. \]