ECE595 / STAT598: Machine Learning I
Lecture 23.1: Probability Inequality - Basic Inequalities

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

Outline

- Lecture 22 Is Learning Feasible?
- Lecture 23 Probability Inequality
- Lecture 24 Probably Approximate Correct

Today's Lecture:

- Basic Inequalities
  - Markov and Chebyshev
  - Interpreting the results
- Advance Inequalities
  - Chernoff inequality
  - Hoeffding inequality

Empirical Average

- We want to take a detour to talk about probability inequalities
- These inequalities will become useful when studying learning theory

Let us look at 1D case.

- You have random variables $X_1, X_2, \ldots, X_N$.
- Assume **independently identically distributed** i.i.d.
- This implies

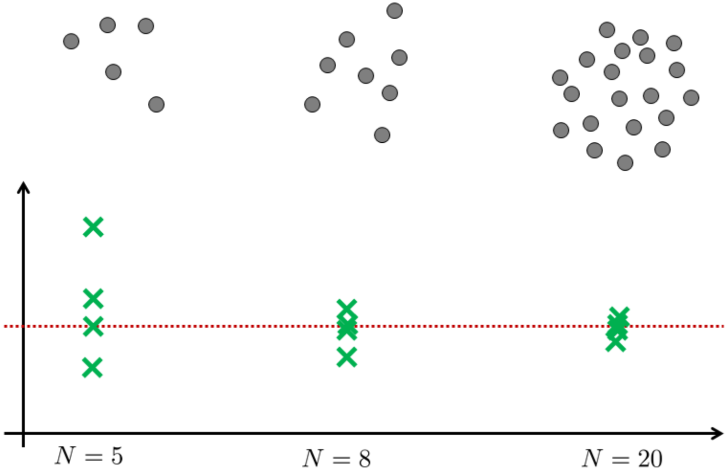
$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = \ldots = \mathbb{E}[X_N] = \mu$$

- You compute the **empirical average**

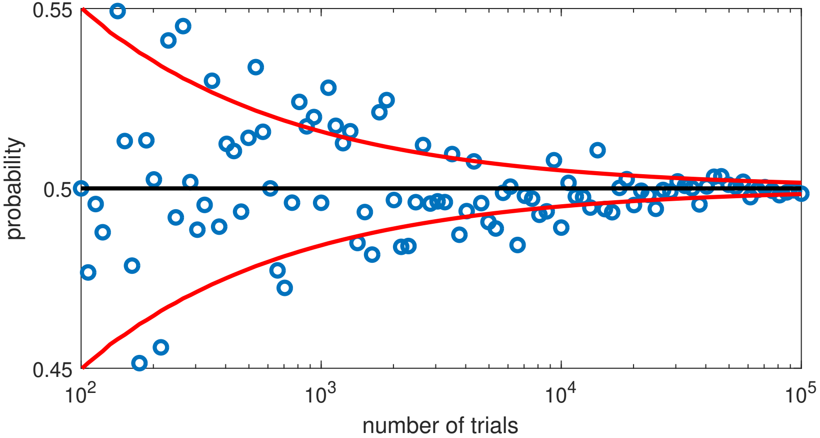
$$\nu = \frac{1}{N} \sum_{n=1}^{N} X_n$$

- How close is ν to μ?

As N grows ...



As N grows ...



Interpreting the Empirical Average

$$\nu = \frac{1}{N}\sum_{n=1}^{N} X_n$$

- ν is a random variable
- ν has CDF and PDF
- ν has mean
-

$$\begin{aligned}\mathbb{E}[\nu] = \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} X_n\right] &= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[X_n] \\ &= \frac{1}{N}N\mu = \mu.\end{aligned}$$

- Note that "$\mathbb{E}[\nu] = \mu$" is not the same as "$\nu = \mu$".
- What is the probability ν deviates from μ?

Probability of Bad Event

$$\mathbb{P}\left[|\nu - \mu| > \epsilon\right] = \ ?$$

- $\mathcal{B} = \{|\nu - \mu| > \epsilon\}$: The $\mathcal{B}$ad event: ν deviates from μ by at least ϵ
- $\mathbb{P}[\mathcal{B}]$ = probability that this bad event happens.
- Want $\mathbb{P}[\mathcal{B}]$ small. So upper bound it by δ.

$$\mathbb{P}\left[|\nu - \mu| > \epsilon\right] \ \leq \ \delta.$$

- With probability **no greater** than δ, $\mathcal{B}$ad event happens.
- Rearrange the equation:

$$\mathbb{P}\left[|\nu - \mu| \leq \epsilon\right] \ > \ 1 - \delta.$$

- With probability **at least** $1 - \delta$, the $\mathcal{B}$ad event will **not** happen.

Markov Inequality

Theorem (Markov Inequality)

For any $X > 0$ *and* $\epsilon > 0$,

$$\mathbb{P}[X \geq \epsilon] \leq \frac{\mathbb{E}[X]}{\epsilon}.$$

$$\begin{aligned}
\epsilon \mathbb{P}[X \geq \epsilon] &= \epsilon \int_{\epsilon}^{\infty} p(x) dx \\
&= \int_{\epsilon}^{\infty} \epsilon \; p(x) dx \\
&\leq \int_{\epsilon}^{\infty} x p(x) dx \\
&\leq \int_{0}^{\infty} x p(x) dx = \mathbb{E}[X].
\end{aligned}$$

Chebyshev Inequality

Theorem (Chebyshev Inequality)

Let $X_1, \ldots, X_N$ *be i.i.d. with* $\mathbb{E}[X_n] = \mu$ *and* $\mathrm{Var}[X_n] = \sigma^2$*. Define*

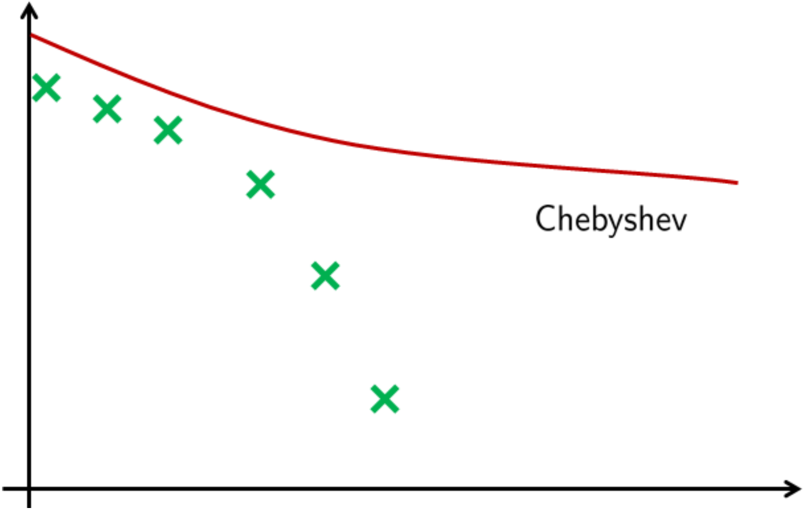
$$\nu = \frac{1}{N} \sum_{n=1}^{N} X_n.$$

Then,

$$\mathbb{P}\left[|\nu - \mu| > \epsilon\right] \le \frac{\sigma^2}{N\epsilon^2}$$

$$\mathbb{P}\left[|\nu - \mu|^2 > \epsilon^2\right] \underbrace{\le \frac{\mathbb{E}[|\nu - \mu|^2]}{\epsilon^2}}_{\text{Markov}} \quad \underbrace{= \frac{\mathrm{Var}[\nu]}{\epsilon^2}}_{\mathbb{E}[(\nu-\mu)^2]=\text{var}[\nu]} \quad \underbrace{= \frac{\sigma^2}{N\epsilon^2}}_{\text{var}[\nu]=\frac{\sigma^2}{N}}.$$

How Good is Chebyshev Inequality?



Weak Law of Large Number

Theorem (WLLN)

Let $X_1, \ldots, X_N$ be a sequence of i.i.d. random variables with common mean μ. Let $M_N = \frac{1}{N}\sum_{n=1}^{N} X_n$. Then, for any $\varepsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}\left[|M_N - \mu| > \varepsilon\right] = 0. \tag{1}$$

Remark:

- The limit is outside the probability.
- This means that the probability of the event $|M_N - \mu| > \varepsilon$ is diminishing as $N \to$.
- But diminishing probability can still have occasions where $|M_N - \mu| > \varepsilon$.
- It just means that these occasions do not happen often.

Strong Law of Large Number

Theorem (SLLN)

Let $X_1, \ldots, X_N$ *be a sequence of i.i.d. random variables with common mean* μ*. Let* $M_N = \frac{1}{N}\sum_{n=1}^{N} X_n$*. Then, for any* $\varepsilon > 0$*,*

$$\mathbb{P}\left[\lim_{N\to\infty} |M_N - \mu| > \varepsilon\right] = 0. \tag{2}$$

Remark:

- The limit is inside the probability.
- We need to analyze the limiting object $\lim_{N\to\infty} |M_N - \mu|$
- This object may or may not exist. This object is another random variable.
- The probability is measuring the event that this limiting object will deviate significantly from ε
- There is no "occasional" outliers.