Markov and Chebyshev Inequalities
Law of Large Numbers
Central Limit Theorem
Markov’s Inequality

**Theorem**

[Markov Inequality] For any non-negative random variable, and for all $a > 0$,

$$\Pr(X \geq a) \leq \frac{E[X]}{a}.$$  

$$E[X] = \int_0^\infty x f_X(x) \, dx \geq \int_a^\infty x f_X(x) \, dx \geq \int_a^\infty a f_X(x) \, dx = a P(X \geq a)$$

\[ y = x \]

\[ f_X(x) \]
Characterizing Random Variables

Suppose we care about some summary of random variable $X$:

$$ Y = g(X) $$

- **Expected Value:** Value of a “typical” sample

\[ E[Y] = E[g(X)] = \int g(x) f_X(x) \, dx \]

- **Variance:** Expected squared distance from mean

\[ \text{Var}[Y] = E[(g(X) - E[g(X)])^2] \]

This is often tractable to compute, and leads to bounds on “extreme” events!

- **Cumulative distribution:** A complete characterization of $g(X)$

\[ F_Y(y) = P(Y \leq y) = P(g(X) \leq y) \]

\[ f_Y(y) = \frac{dF_Y(y)}{dy} \]

If $g(X)$ is complex or high-dimensional, may be hard to compute!
Theorem

For any random variable $X$, and any $a > 0$,

$$ Pr(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}. $$

Proof.

$$ Pr(|X - E[X]| \geq a) = Pr((X - E[X])^2 \geq a^2) $$

By Markov inequality

$$ Pr((X - E[X])^2 \geq a^2) \leq \frac{E[(X - E[X])^2]}{a^2} = \frac{Var[X]}{a^2} $$
Chebyshev’s Inequality

Theorem

For any random variable $X$, and any $a > 0$,

$$Pr(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}.$$ 

- Another way of parameterizing Chebyshev’s inequality:

$$\mu = E[X], \quad \sigma = \sqrt{Var[X]}$$

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- Chebyshev bound is vacuous (above one) for events less than one standard deviation from the mean. But this could be likely!
Another way of parameterizing Chebyshev’s inequality:

\[ \mu = E[X], \quad \sigma = \sqrt{\text{Var}[X]} \]

\[ P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \]

Chebyshev bound is vacuous (above one) for events less than one standard deviation from the mean. But this could be likely!
Markov and Chebyshev Inequalities
Law of Large Numbers
Central Limit Theorem
Convergence in Probability

Convergence in Probability

Let $Y_1, Y_2, \ldots$ be a sequence of random variables (not necessarily independent), and let $a$ be a real number. We say that the sequence $Y_n$ converges to $a$ in probability, if for every $\epsilon > 0$, we have

$$\lim_{n \to \infty} P(|Y_n - a| \geq \epsilon) = 0.$$

“(almost all) of the PMF/PDF of $Y_n$ eventually gets concentrated (arbitrarily) close to $a”

Convergence of a Deterministic Sequence

Let $a_1, a_2, \ldots$ be a sequence of real numbers, and let $a$ be another real number. We say that the sequence $a_n$ converges to $a$, or $\lim_{n \to \infty} a_n = a$, if for every $\epsilon > 0$ there exists some $n_0$ such that

$$|a_n - a| \leq \epsilon, \quad \text{for all } n \geq n_0.$$

“$a_n$ eventually gets and stays (arbitrarily) close to $a”
Convergence in Probability

Convergence in Probability

Let $Y_1, Y_2, \ldots$ be a sequence of random variables (not necessarily independent), and let $a$ be a real number. We say that the sequence $Y_n$ converges to $a$ in probability, if for every $\epsilon > 0$, we have

$$\lim_{n \to \infty} P(|Y_n - a| \geq \epsilon) = 0.$$ 

Example:

$X_n$ is a sequence of independent uniform variables on $[0, 1]$ and $Y_n = \min\{X_1, \ldots, X_n\}$

We expect that $Y_n$ converges to zero. To verify:

$$P \left( |Y_n - 0| \geq \epsilon \right) = P(X_1 \geq \epsilon, \ldots, X_n \geq \epsilon)$$

$$= P(X_1 \geq \epsilon) \cdots P(X_n \geq \epsilon)$$

$$= (1 - \epsilon)^n.$$
7.2 The Weak Law of Large Numbers

Let \( X_1, X_2, \ldots \) i.i.d. with finite mean \( \mu \) and variance \( \sigma^2 \).

The sample mean is defined by

\[
M_n = \frac{X_1 + \cdots + X_n}{n}
\]

\( M_n \) is the empirical mean.

\[
E[M_n] = \frac{E[X_1] + \cdots + E[X_n]}{n} = \frac{n\mu}{n} = \mu,
\]

\[
\text{Var}[M_n] = \frac{\text{Var}(X_1 + \cdots + X_n)}{n^2} = \frac{\text{Var}(X_1) + \cdots + \text{Var}(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.
\]

Chebyshev’s inequality bounds distance between the true mean and the “empirical” or “sample” mean:

\[
P(\left| M_n - \mu \right| \geq \epsilon) \leq \frac{\text{Var}(M_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}.
\]

The empirical mean converges to the true mean in probability

\[
\lim_{n \to \infty} P(\left| M_n - \mu \right| \geq \epsilon) = 0
\]

True even if variance not finite, but proof more challenging.
Why is it a “Weak” Law of Large Numbers?

Convergence in Probability
Let $Y_1, Y_2, \ldots$ be a sequence of random variables (not necessarily independent), and let $a$ be a real number. We say that the sequence $Y_n$ converges to $a$ in probability, if for every $\epsilon > 0$, we have

$$\lim_{n \to \infty} P(|Y_n - a| \geq \epsilon) = 0.$$ 

Example:

For every $\epsilon > 0$, $\lim_{n \to \infty} P(|Y_n - 0| \geq \epsilon) = 0$.

But even though $Y_n$ converges in probability, occasionally it takes on very large values:

$$E[Y_n] = 1 \text{ for all } n.$$
The Strong Law of Large Numbers (SLLN)

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with mean $\mu$. Then, the sequence of sample means $M_n = (X_1 + \cdots + X_n)/n$ converges to $\mu$, with probability 1, in the sense that

$$P\left( \lim_{n \to \infty} \frac{X_1 + \cdots + X_n}{n} = \mu \right) = 1.$$ 

- This stronger (but more technically challenging) notion of convergence rules out cases like the previous example.
- For many practical scenarios, both forms of convergence hold, but convergence in probability is easier to show.
- We focus exclusively on the weak law in this course.
CS145: Lecture 15 Outline

- Markov and Chebyshev Inequalities
- Law of Large Numbers
- Central Limit Theorem
Scaling of the Sample Mean

- Sequence of *independent, identically distributed* random variables:

  \[ X_1, X_2, \ldots, X_n \quad E[X_i] = \mu \quad \text{Var}[X_i] = \sigma^2 < \infty \]

- The variance of their *sum* increases with \( n \):

  \[ S_n = \sum_{i=1}^{n} X_i \quad E[S_n] = n\mu \quad \text{Var}[S_n] = n\sigma^2 \]

- **Law of Large Numbers**: variance of the *empirical mean* decreases with \( n \):

  \[ M_n = \frac{1}{n} S_n \quad E[M_n] = \mu \quad \text{Var}[M_n] = \frac{\sigma^2}{n} \]

- **Standardized sum**: transform so mean and variance constant for all \( n \)

  \[ Z_n = \frac{S_n - E[S_n]}{\sqrt{\text{Var}[S_n]}} = \frac{S_n - n\mu}{\sqrt{n\sigma}} \quad E[Z_n] = 0 \quad \text{Var}[Z_n] = 1 \]

What is the shape of the distribution of \( Z_n \) for large \( n \)?
Central Limit Theorem (CLT)

- “Standardized” $S_n = X_1 + \cdots + X_n$:
  \[ Z_n = \frac{S_n - \mathbb{E}[S_n]}{\sigma_{S_n}} = \frac{S_n - n\mathbb{E}[X]}{\sqrt{n}\sigma} \]
  - zero mean
  - unit variance

- Let $Z$ be a standard normal r.v. (zero mean, unit variance)
  \[ f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \]

- **Theorem:** For every $c$:
  \[ P(Z_n \leq c) \rightarrow P(Z \leq c) \]

- $P(Z \leq c)$ is the standard normal CDF, $\Phi(c)$, available from the normal tables
Central Limit Theorem (CLT)

Usefulness
- universal; only means, variances matter
- accurate computational shortcut
- justification of normal models

What exactly does it say?
- CDF of \( Z_n \) converges to normal CDF
  - not a statement about convergence of PDFs or PMFs
- Treat \( Z_n \) as if normal
  - also treat \( S_n \) as if normal

Can we use it when \( n \) is “moderate”?
- Yes, but no nice theorems to this effect
- Symmetry helps a lot

Theorem: For every \( c \):
\[
P(Z_n \leq c) \to P(Z \leq c)
\]

\( P(Z \leq c) \) is the standard normal CDF, \( \Phi(c) \), available from the normal tables
$S_n = \sum_{i=1}^{n} X_i$

$f_{X_i}(x_i) = 1$ if $0 \leq x_i \leq 1$, 0 otherwise.
CLT: Exponential Random Variables

\[ f_{X_i}(x) = \lambda e^{-\lambda x}, \quad x \geq 0. \]

\[ E[X_i] = \frac{1}{\lambda} \quad \text{Var}[X_i] = \frac{1}{\lambda^2} \]

\[ S_n = \sum_{i=1}^{n} X_i \]
**Theorem:** A sum of independent Gaussians is Gaussian

\[ S = \sum_{i=1}^{n} X_i \]

\[ f_S(s) = \frac{1}{\sqrt{2\pi \bar{\sigma}^2}} e^{-\frac{1}{2} \left( \frac{s - \bar{\mu}}{\bar{\sigma}} \right)^2} \]

\[ f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi \sigma_i^2}} e^{-\frac{1}{2} \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2} \]

\[ \bar{\mu} = \sum_{i=1}^{n} \mu_i, \quad \bar{\sigma}^2 = \sum_{i=1}^{n} \sigma_i^2 \]

A family of distributions is **stable** if a linear combination of independent random variables stays in the same family

**Theorem:** The only (non-degenerate) stable family of random variables with finite variance is the Gaussian.

**One proof:** Take Taylor expansion of log-pdf (or its Fourier transform), show that all terms of order three or higher approach zero as \( n \) grows, giving a quadratic (Gaussian) limit
CLT: Binomial Distribution

- Fix $p$, where $0 < p < 1$
- $X_i$: Bernoulli($p$)
- $S_n = X_1 + \cdots + X_n$: Binomial($n, p$)
  - mean $np$, variance $np(1-p)$
- CDF of $\frac{S_n - np}{\sqrt{np(1-p)}} \rightarrow$ standard normal

De Moivre – Laplace Approximation to the Binomial

If $S_n$ is a binomial random variable with parameters $n$ and $p$, $n$ is large, and $k, \ell$ are nonnegative integers, then

$$P(k \leq S_n \leq \ell) \approx \Phi \left( \frac{\ell + \frac{1}{2} - np}{\sqrt{np(1-p)}} \right) - \Phi \left( \frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}} \right).$$

- $P(S_n \leq 21) = P(S_n < 22)$, because $S_n$ is integer
- Compromise: consider $P(S_n \leq 21.5)$
When the 1/2 correction is used, CLT can also approximate the binomial p.m.f. (not just the binomial CDF)

\[
P(S_n = 19) = P(18.5 \leq S_n \leq 19.5) \\
n = 36, p = 0.5 \\
18.5 \leq S_n \leq 19.5 \\
18.5 - 18 \leq S_n - 18 \leq 19.5 - 18 \\
\frac{18.5 - 18}{3} \leq \frac{S_n - 18}{3} \leq \frac{19.5 - 18}{3} \\
0.17 \leq Z_n \leq 0.5 \\
P(S_n = 19) \approx P(0.17 \leq Z \leq 0.5) \\
= P(Z \leq 0.5) - P(Z \leq 0.17) \\
= 0.6915 - 0.5675 \\
= 0.124
\]

- Exact answer:

\[
\binom{36}{19} \left(\frac{1}{2}\right)^{36} = 0.1251
\]
Pollster’s Problem: Chebyshev

- \( f \): fraction of population that “...”
- \( i \)th (randomly selected) person polled:
  \[
  X_i = \begin{cases} 
    1, & \text{if yes,} \\
    0, & \text{if no.} 
  \end{cases}
  \]
- Use Chebyshev’s inequality:
  \[
  P(\left| M_n - f \right| \geq .01) \leq \frac{\sigma_{M_n}^2}{(0.01)^2} = \frac{\sigma_x^2}{n(0.01)^2} \leq \frac{1}{4n(0.01)^2}
  \]
- If \( n = 50,000 \), then \( P(\left| M_n - f \right| \geq .01) \leq .05 \) (conservative)
- \( M_n = (X_1 + \cdots + X_n)/n \) fraction of “yes” in our sample
- Goal: 95% confidence of \( \leq 1\% \) error
  \[
  P(\left| M_n - f \right| \geq .01) \leq .05
  \]

**For any binary variable,**

\[
\text{Var}(X_i) \leq \frac{1}{2^2}
\]
Pollster’s Problem: CLT

- $f$: fraction of population that “…”
- $i$th (randomly selected) person polled:
  \[ X_i = \begin{cases} 
  1, & \text{if yes,} \\
  0, & \text{if no.} 
  \end{cases} \]
- Event of interest: $|M_n - f| \geq .01$
  \[ \left| \frac{X_1 + \cdots + X_n - nf}{n} \right| \geq .01 \]
  \[ \left| \frac{X_1 + \cdots + X_n - nf}{\sqrt{n}\sigma} \right| \geq \frac{.01\sqrt{n}}{\sigma} \]
  \[ P(|M_n - f| \geq .01) \approx P(|Z| \geq .01\sqrt{n}/\sigma) \]
  \[ \leq P(|Z| \geq .02\sqrt{n}) \]
  \[ \sqrt{n} = \frac{2}{0.02} = 100, \quad n = 10,000 \]
- $M_n = (X_1 + \cdots + X_n)/n$ fraction of “yes” in our sample
- Goal: 95% confidence of $\leq 1\%$ error
  \[ P(|M_n - f| \geq .01) \leq .05 \]

For any binary variable,
\[ \text{Var}(X_i) \leq \frac{1}{2^2} \quad \text{Std}(X_i) \leq \frac{1}{2} \]