Weak Law of Large Numbers

• Consider i.i.d. random variables $X_1, X_2, \ldots$
  
  ▪ $X_i$ have distribution $F$ with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$
  
  ▪ Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
  
  ▪ For any $\varepsilon > 0$:
    
    $P(\bar{X} - \mu \geq \varepsilon) \xrightarrow{n \to \infty} 0$

• Proof:

  $E[\bar{X}] = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] = \mu \quad \text{Var}(\bar{X}) = \text{Var}\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}$

  ▪ By Chebyshev’s inequality:

    $P(\left|\bar{X} - \mu\right| \geq \varepsilon) \leq \frac{\sigma^2}{n \varepsilon^2} \xrightarrow{n \to \infty} 0$
Strong Law of Large Numbers

- Consider I.I.D. random variables $X_1, X_2, \ldots$
  - $X_i$ have distribution $F$ with $E[X_i] = \mu$
  - Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
  - $P\left(\lim_{n \to \infty} \left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \mu\right) = 1$

- Strong Law $\Rightarrow$ Weak Law, but not vice versa
- Strong Law implies that for any $\varepsilon > 0$, there are only a finite number of values of $n$ such that condition of Weak Law: $|\bar{X} - \mu| \geq \varepsilon$ holds.
Intuitions and Misconceptions of LLN

• Say we have repeated trials of an experiment
  - Let event E = some outcome of experiment
  - Let $X_i = 1$ if E occurs on trial $i$, 0 otherwise
  - Strong Law of Large Numbers (Strong LLN) yields:
    \[
    \frac{X_1 + X_2 + \ldots + X_n}{n} \to E[X_i] = P(E)
    \]
  - Recall first week of class: $P(E) = \lim_{n \to \infty} n(E)/n$
  - Strong LLN justifies “frequency” notion of probability
  - Misconception arising from LLN:
    - Gambler’s fallacy: “I’m due for a win”
    - Consider being “due for a win” with repeated coin flips...
La Loi des Grands Nombres

- **History of the Law of Large Numbers**
  - 1713: Weak LLN described by Jacob Bernoulli
  - 1835: Poisson calls it “La Loi des Grands Nombres”
    - That would be “Law of Large Numbers” in French
  - 1909: Émile Borel develops Strong LLN for Bernoulli random variables
  - 1928: Andrei Nikolaevich Kolmogorov proves Strong LLN in general case
And now a moment of silence...

...before we present...

...the greatest result of probability theory!
The Central Limit Theorem (CLT)

- Consider I.I.D. random variables $X_1, X_2, ...$
  - $X_i$ have distribution $F$ with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$
  \[
  \frac{X_1 + X_2 + ... + X_n - n\mu}{\sigma\sqrt{n}} \rightarrow N(0, 1) \quad \text{as} \quad n \rightarrow \infty
  \]
- More intuitively:
  - Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
  - Central Limit Theorem: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ \quad \text{as} \quad n \rightarrow \infty$
  - Now let $Z = \frac{\bar{X} - \mu}{\sigma \sqrt{n}}$, noting that $Z \sim N(0, 1)$:
    \[
    \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \iff Z = \frac{1}{n} \left( \sum_{i=1}^{n} X_i \right) - \mu = n \left[ \frac{1}{n} \left( \sum_{i=1}^{n} X_i \right) - \mu \right] = \frac{1}{n} \left( n \sum_{i=1}^{n} X_i \right) - n\mu = \left( \sum_{i=1}^{n} X_i \right) - n\mu
    \]
No Limits for Central Limit Theorem

- History of the Central Limit Theorem
  - 1733: CLT for $X \sim \text{Ber}(1/2)$ postulated by Abraham de Moivre
  - 1823: Pierre-Simon Laplace extends de Moivre’s work to approximating $\text{Bin}(n, p)$ with Normal
  - 1901: Aleksandr Lyapunov provides precise definition and rigorous proof of CLT
  - 2003: Charlie Sheen stars in television series “Two and a Half Men”
    - By end of the 7th (final) season, there were 161 episodes
    - Mean quality of subsamples of episodes is Normally distributed (thanks to the Central Limit Theorem)
Central Limit Theorem in Real World

- CLT is why many things in “real world” appear Normally distributed
  - Many quantities are sum of independent variables
  - Exams scores
    - Sum of individual problems
  - Election polling
    - Ask 100 people if they will vote for candidate X \((p_1 = \# \text{“yes”}/100)\)
    - Repeat this process with different groups to get \(p_1, \ldots, p_n\)
    - Will have a normal distribution over \(p_i\)
    - Can produce a “confidence interval”
      - How likely is it that estimate for true \(p\) is correct
      - We’ll do an example like that soon
A Prior CS109 Midterm on the CLT

- Start with 370 midterm scores: $X_1, X_2, \ldots, X_{370}$
  - $E[X_i] = 79.5$ and $\text{Var}(X_i) = 417.87$
  - Created 37 disjoint samples of size $n = 10$
    - $Y_1 = \{X_1, \ldots, X_{10}\}, Y_2 = \{X_{11}, \ldots, X_{20}\}, \ldots, Y_{37} = \{X_{100}, \ldots, X_{370}\}$
  - Prediction by CLT: $\frac{\bar{Y}_i}{10} = \frac{\sum_{j=10i}^{10i+9} X_j}{10} \\ \bar{Y}_i \sim N(79.5, 417.87/10 \approx 41.787)$

$$Z_i = \frac{\bar{Y}_i - E[X_i]}{\sqrt{\sigma^2/n}} = \frac{\bar{Y}_i - 79.5}{\sqrt{417.87/10}} \quad \bar{Z} = \frac{1}{37} \sum_{i=1}^{37} Z_i = 4.74 \times 10^{-16} \quad \text{Var}(Z_i) = 0.96$$

![Histogram of Z_i values]
Estimating Clock Running Time

- Have new algorithm to test for running time
  - Mean (clock) running time: $\mu = t$ sec.
  - Variance of running time: $\sigma^2 = 4$ sec$^2$.
  - Run algorithm repeatedly (I.I.D. trials), measure time
    - How many trials so estimated time = $t \pm 0.5$ with 95% certainty?
    - $X_i$ = running time of $i$-th run (for $1 \leq i \leq n$)
    - By Central Limit Theorem, $Z \sim N(0, 1)$, where:
      $$Z_n = \frac{\left( \sum_{i=1}^{n} X_i \right) - n\mu}{\sigma \sqrt{n}} = \frac{\left( \sum_{i=1}^{n} X_i \right) - nt}{2 \sqrt{n}}$$
      $$P(-0.5 \leq \frac{\sum_{i=1}^{n} X_i}{n} - t \leq 0.5) = P\left(-0.5 \sqrt{n} \leq \frac{\sqrt{n} \left( \sum_{i=1}^{n} X_i \right) - nt}{2 \sqrt{n}} \leq 0.5 \sqrt{n}\right) = P\left(-0.5 \sqrt{n} \leq Z_n \leq 0.5 \sqrt{n}\right)$$
      $$= \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = \Phi\left(\frac{\sqrt{n}}{4}\right) - (1 - \Phi\left(\frac{\sqrt{n}}{4}\right)) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \approx 0.95 \quad \Rightarrow \quad \Phi\left(\frac{\sqrt{n^*}}{4}\right) = 0.975$$
      - Solve for $n^*$: $\frac{\sqrt{n^*}}{4} = 1.96$ \quad $\Rightarrow \quad n^* = \left\lceil (7.84)^2 \right\rceil = 62$
Estimating Time With Chebyshev

- Have new algorithm to test for running time
  - Mean (clock) running time: \( \mu = t \) sec.
  - Variance of running time: \( \sigma^2 = 4 \) sec\(^2\).
  - Run algorithm repeatedly (I.I.D. trials), measure time
    - How many trials so estimated time = \( t \pm 0.5 \) with 95% certainty?
    - \( X_i = \) running time of \( i \)-th run (for \( 1 \leq i \leq n \)), and \( X_S = \sum_{i=1}^{n} \frac{X_i}{n} \)

- What would Chebyshev say?
  \[
P\left(\left| X_S - \mu_S \right| \geq k \right) \leq \frac{\sigma_S^2}{k^2}
\]

\[
\mu_S = E\left[ \sum_{i=1}^{n} \frac{X_i}{n} \right] = t \quad \sigma_S^2 = \text{Var}\left( \sum_{i=1}^{n} \frac{X_i}{n} \right) = \sum_{i=1}^{n} \text{Var}\left( \frac{X_i}{n} \right) = n \frac{\sigma^2}{n^2} = \frac{4}{n}
\]

\[
P\left( \left| \sum_{i=1}^{n} \frac{X_i}{n} - t \right| \geq 0.5 \right) \leq \frac{4/n}{(0.5)^2} = \frac{16}{n} = 0.05 \quad \Rightarrow \quad n \geq 320
\]

Thanks for playing, Pafnuty!
Crashing Your Web Site

- Number visitors to web site/minute: \( X \sim \text{Poi}(100) \)
  - Server crashes if \( \geq 120 \) requests/minute
  - What is \( P(\text{crash in next minute})? \)
  - Exact solution: \( P(X \geq 120) = \sum_{i=120}^{\infty} \frac{e^{-100} (100)^i}{i!} \approx 0.0282 \)
  - Use CLT, where \( \text{Poi}(100) \sim \sum_{i=1}^{n} \text{Poi}(100/n) \) (all I.I.D)

\[
P(X \geq 120) = P(Y \geq 119.5) = P\left( \frac{Y - 100}{\sqrt{100}} \geq \frac{119.5 - 100}{\sqrt{100}} \right) = 1 - \Phi(1.95) \approx 0.0256
\]

- Note: Normal can be used to approximate Poisson
  - I’ll give you one more chance (one-sided) Chebyshev:

\[
P(X \geq 120) = P(X \geq E[X] + a) \leq \frac{\sigma^2}{\sigma^2 + a^2} = \frac{100}{100 + 20^2} = 0.2
\]
It’s play time!
Sum of Dice

• You will roll 10 6-sided dice \((X_1, X_2, \ldots, X_{10})\)
  - \(X = \text{total value of all 10 dice} = X_1 + X_2 + \ldots + X_{10}\)
  - Win if: \(X \leq 25\) or \(X \geq 45\)
  - Roll!

• And now the truth (according to the CLT)…
Sum of Dice

- You will roll 10 6-sided dice \((X_1, X_2, \ldots, X_{10})\)
  - \(X = \text{total value of all 10 dice} = X_1 + X_2 + \ldots + X_{10}\)
  - Win if: \(X \leq 25\) or \(X \geq 45\)

- Recall CLT: 
  \[
  \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma \sqrt{n}} \rightarrow N(0,1) \quad \text{as} \quad n \rightarrow \infty
  \]
  
  - Determine \(P(X \leq 25 \text{ or } X \geq 45)\) using CLT:
  \[
  \mu = E[X_i] = 3.5 \quad \quad \sigma^2 = \text{Var}(X_i) = \frac{35}{12}
  \]
  
  
  \[
  1 - P(25.5 \leq X \leq 44.5) = 1 - P\left(\frac{25.5 - 10(3.5)}{\sqrt{35/12}/\sqrt{10}} \leq \frac{X - 10(3.5)}{\sqrt{35/12}/\sqrt{10}} \leq \frac{44.5 - 10(3.5)}{\sqrt{35/12}/\sqrt{10}}\right)
  \]
  
  \[
  \approx 1 - (2\Phi(1.76) - 1) \approx 2(1 - 0.9608) = 0.0784
  \]