Weak Law of Large Numbers

- Consider I.I.D. random variables \( X_1, X_2, \ldots \)
  - \( X_i \) have distribution \( F \) with \( \text{E}[X_i] = \mu \) and \( \text{Var}(X_i) = \sigma^2 \)
  - Let \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \)
  - For any \( \varepsilon > 0 \):
    \[
    P\left( \left| \bar{X} - \mu \right| \geq \varepsilon \right) \rightarrow 0
    \]

  - Proof:
    \[
    \text{E}[\bar{X}] = \text{E}\left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n} \sum_{i=1}^{n} \text{E}[X_i] = \mu
    \]
    \[
    \text{Var}(\bar{X}) = \text{Var}\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{\sigma^2}{n}
    \]
  - By Chebyshev’s inequality:
    \[
    P\left( \left| \bar{X} - \mu \right| \geq \varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0
    \]

Strong Law of Large Numbers

- Consider I.I.D. random variables \( X_1, X_2, \ldots \)
  - \( X_i \) have distribution \( F \) with \( \text{E}[X_i] = \mu \)
  - Let \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \)
    \[
    \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu
    \]
  - 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: \( \left| \bar{X} - \mu \right| \geq \varepsilon \) holds.

Intuitions and Misconceptions of LLN

- Say we have repeated trials of an experiment
  - Let event \( E \) be some outcome of experiment
    - \( E \) occurs on trial \( i \), 0 otherwise
  - Strong Law of Large Numbers (Strong LLN) yields:
    \[
    \frac{X_1 + X_2 + \ldots + X_n}{n} \rightarrow \text{E}[X] = P(E)
    \]
  - Recall first week of class: \( P(E) = \lim_{n \rightarrow \infty} \frac{m(E)}{n} \)
  - 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
  - 2011: Another year passes in which Charlie Sheen does not make use of LLN
    - I’m still holding out hope for 2012...

The Central Limit Theorem (CLT)

- Consider I.I.D. random variables \( X_1, X_2, \ldots \)
  - \( X_i \) have distribution \( F \) with \( \text{E}[X_i] = \mu \) and \( \text{Var}(X_i) = \sigma^2 \)
    \[
    \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \rightarrow N(0, 1)
    \]
  - As \( n \rightarrow \infty \)
  - More intuitively:
    - Let \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \)
      \[
      \bar{X} \sim N\left( \mu, \frac{\sigma^2}{n} \right)
      \]
    - Central Limit Theorem: \( \bar{X} \sim N\left( \mu, \frac{\sigma^2}{n} \right) \) as \( n \rightarrow \infty \)
    - Now let \( Z = \frac{\bar{X} - \mu}{\sigma\sqrt{n}} \)
      \[
      \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \Rightarrow Z \sim N(0, 1)
      \]
    - \( Z \sim N(0, 1) \)
      \[
      Z = \frac{\bar{X} - \mu}{\sigma\sqrt{n}} = \frac{n\left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] - \mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}
      \]

Silence!!

And now a moment of silence...

...before we present...

...the greatest result of probability theory!
No Limits for Central Limit Theorem

- History of the Central Limit Theorem
  - 1733: CLT for X ~ Ber(1/2) postulated by Abraham de Moivre
  - 1823: Pierre-Simon Laplace extends de Moivre’s work to approximating 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 Half Men”
    - By end of the 7th season (last year), 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 (p1 = # “yes”/100)
  - Repeat this process with different groups to get p1, ..., pn
  - Have a normal distribution over pi
  - Can produce a “confidence interval”
    - How likely is it that estimate for true p is correct
    - We’ll do an example like that soon

This is Your Midterm on the CLT

- Start with 180 midterm scores: X1, X2, ..., X180
  - E[X] = 68.9 and Var(X) = 611.37
  - Created 18 disjoint samples of size n = 10
    - Y1 = (X1, ..., X10), Y2 = (X11, ..., X20), Yn = (X101, ..., X110)
  - Prediction by CLT:
    - \( Z = \frac{\overline{Y} - E[Y]}{\sigma/\sqrt{n}} \)
    - \( Z \sim N(0, 1) \) where:
      - \( \overline{Y} = \frac{\sum X_i}{n} \)
      - \( \sigma^2 = \frac{\sum (X_i - \overline{X})^2}{n} \)
      - \( \sigma = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} \)
- By Central Limit Theorem, Z ~ N(0, 1), where:
  - \( Z = n \left( \sum_{i=1}^{n} X_i - n \mu \right) \sqrt{n} \)
  - \( Z \sim N(0, 1) \)
  - \( \mu = E[X] \)
  - \( \sigma^2 = Var(X) \)
- CLT is why many things in “real world” appear Normally distributed

Estimating Clock Running Time

- Have new algorithm to test for running time
  - Mean (clock) running time: \( \mu = t \) sec.
  - Variance of running time: \( \sigma^2 = 4 \) sec²
- Run algorithm repeatedly (I.I.D. trials), measure time
  - How many trials so estimated time = \( t \pm 0.5 \) with 95% certainty?
  - \( X_i \sim E[X_i] = t \)
  - \( Y = \frac{1}{n} \sum X_i \)
  - \( \bar{y} = \frac{1}{n} \sum X_i \)
  - \( \bar{Y} = \frac{1}{n} \sum Y_i \)
  - \( Z = \frac{\bar{Y} - E[Y]}{\sigma/\sqrt{n}} \)
  - \( Z \sim N(0, 1) \)
  - By Central Limit Theorem, Z ~ N(0, 1), where:
    - \( Z = n \left( \sum_{i=1}^{n} X_i - n \mu \right) \sqrt{n} \)
    - \( Z \sim N(0, 1) \)
    - \( \mu = E[X] \)
    - \( \sigma^2 = Var(X) \)
- What would Chebyshev say?
  - \( P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2} \)
  - \( \mu = \frac{\sum X_i}{n} \)
  - \( \sigma^2 = \frac{\sum (X_i - \overline{X})^2}{n} \)
  - \( \sigma = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} \)
  - \( P\left( \left| \frac{\sum X_i}{n} - t \right| \geq 0.5 \right) \leq \frac{4/n}{0.5} = 16/n = 0.05 \Rightarrow n \geq 320 \)
  - Thanks for playing Pafnuty...

Estimating Time With Chebyshev

- Have new algorithm to test for running time
  - Mean (clock) running time: \( \mu = t \) sec.
  - Variance of running time: \( \sigma^2 = 4 \) sec²
- Run algorithm repeatedly (I.I.D. trials), measure time
  - How many trials so estimated time = \( t \pm 0.5 \) with 95% certainty?
  - \( X_i \sim E[X_i] = t \)
  - \( Y = \frac{1}{n} \sum X_i \)
  - \( \bar{y} = \frac{1}{n} \sum X_i \)
  - \( \bar{Y} = \frac{1}{n} \sum Y_i \)
  - \( Z = \frac{\bar{Y} - E[Y]}{\sigma/\sqrt{n}} \)
  - \( Z \sim N(0, 1) \)
  - By Central Limit Theorem, Z ~ N(0, 1), where:
    - \( Z = n \left( \sum_{i=1}^{n} X_i - n \mu \right) \sqrt{n} \)
    - \( Z \sim N(0, 1) \)
    - \( \mu = E[X] \)
    - \( \sigma^2 = Var(X) \)
- What would Chebyshev say?
  - \( P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2} \)
  - \( \mu = \frac{\sum X_i}{n} \)
  - \( \sigma^2 = \frac{\sum (X_i - \overline{X})^2}{n} \)
  - \( \sigma = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} \)
  - \( P\left( \left| \frac{\sum X_i}{n} - t \right| \geq 0.5 \right) \leq \frac{4/n}{0.5} = 16/n = 0.05 \Rightarrow n \geq 320 \)
  - Thanks for playing Pafnuty...

Crashing Your Web Site

- Number visitors to web site/minute: \( X \sim Poi(100) \)
  - Server crashes if \( \geq 120 \) requests/minute
  - What is P(crash in next minute)?
  - Exact solution: \( P(X \geq 120) = \sum_{i=120}^{\infty} e^{-100}(100/r!) \approx 0.0282 \)
  - Use CLT, where Poi(100) = \( \sum_{i=1}^{n} Poi(100/n) \) (all I.I.D)
  - \( P(X \geq 120) = P(X \geq 119.5) \approx 1 - \Phi(1.95) = 0.0256 \)
  - Note: Normal can be approximated to Poisson
  - I’ll give you one more chance (one-sided) Chebyshev:
    - \( P(X \geq 120) = P(X \geq t \cdot E[X] + \sigma) \leq \frac{\sigma^2}{\sigma^2 + \alpha^2} \approx \frac{100}{100 + 20^2} = 0.2 \)
It's play time!

**Sum of Dice**

- You will roll 10 6-sided dice ($X_1, X_2, \ldots, X_{10}$)
  - $X$ = 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):
  
  \[
  E[X] = 10E[X_i] = 10(3.5) = 35
  \]
  
  \[
  \text{Var}(X) = 10 \text{Var}(X_i) = 10 \cdot \frac{350}{12} = \frac{350}{12}
  \]

  \[
  1 - P(25.5 \leq X \leq 44.5) = 1 - P \left( \frac{25.5 - 35}{\sqrt{\frac{350}{12}}} \leq \frac{X - 35}{\sqrt{\frac{350}{12}}} \leq \frac{44.5 - 35}{\sqrt{\frac{350}{12}}} \right)
  \]

  \[
  = 1 - 2(\Phi(1.76) - 1) \approx 2(1 - 0.9332) = 0.0784
  \]

- If only Chebyshev were right...

  \[
  P(|X - \mu| \geq k) = P(|X - 35| \geq 10) \leq \frac{\sigma^2}{k^2} = \frac{350/12}{100} = 0.292
  \]