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 $Var(X_i) = \sigma^2$
  - Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - For any $\varepsilon > 0$:

$$P(|\overline{X} - \mu| \geq \varepsilon) \xrightarrow{n\to\infty} 0$$

- Proof:

$$E[\overline{X}] = E\left[\tfrac{X_1+X_2+\ldots+X_n}{n}\right] = \mu \qquad \mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\tfrac{X_1+X_2+\ldots+X_n}{n}\right) = \frac{\sigma^2}{n}$$

  - By Chebyshev's inequality:

$$P(|\overline{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n\to\infty} 0$$

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 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$P\left(\lim_{n\to\infty}\left(\frac{X_1+X_2+\ldots+X_n}{n}\right) = \mu\right) = 1$$

  - 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: $|\overline{X} - \mu| \geq \varepsilon$ 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:

$$\frac{X_1+X_2+\ldots+X_n}{n} \to E[X] = P(E)$$

  - Recall first week of class: $P(E) = \lim_{n\to\infty}\frac{n(E)}{n}$
  - 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

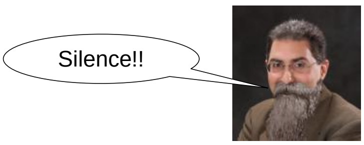
- History of the Law of Large Numbers
  - 1713: Weak LLN described by Jacob Bernoulli

  - 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

  - 2011: Another year passes in which Charlie Sheen does not make use of LLN
    - I'm still holding out hope for 2012...



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 $X_1, X_2, \ldots$
  - X_i have distribution F with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$

$$\frac{X_1+X_2+\ldots+X_n - n\mu}{\sigma\sqrt{n}} \to N(0,1) \quad \text{as} \quad n\to\infty$$

  - More intuitively:

Demo

    - Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
    - Central Limit Theorem: $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ as $n \to \infty$
    - Now let $Z = \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}$, noting that Z ~ N(0, 1):

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \Leftrightarrow Z = \frac{\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right) - \mu}{\sqrt{\sigma^2/n}} = \frac{n\left[\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right) - \mu\right]}{n\sqrt{\sigma^2/n}} = \frac{\left(\sum_{i=1}^{n} X_i\right) - n\mu}{\sigma\sqrt{n}}$$

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 (p_1 = # "yes"/100)
    - Repeat this process with different groups to get $p_1, \ldots, p_n$
    - 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

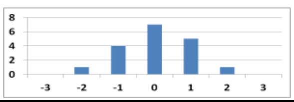
This is Your Midterm on the CLT

- Start with 180 midterm scores: $X_1, X_2, \ldots, X_{180}$
  - $E[X_i] = 68.9$ and $\text{Var}(X_i) = 611.37$
  - Created 18 disjoint samples of size $n = 10$
    - $Y_1 = \{X_1,\ldots, X_{10}\}$, $Y_2 = \{X_{11},\ldots, X_{20}\}$, $Y_i = \{X_{10i-9},\ldots, X_{10i}\}$

$$\bar{Y}_i = \frac{1}{10}\sum_{j=10i-9}^{10i} Y_j$$

  - Prediction by CLT: $\bar{Y}_i \sim \text{N}(68.9,\ 611.37/18 \approx 33.965)$

$$Z_i = \frac{\bar{Y}_i - E[X_i]}{\sqrt{\sigma^2/n}} = \frac{\bar{Y}_i - 68.9}{\sqrt{611.37/18}} \qquad \bar{Z} = \frac{1}{18}\sum_{i=1}^{18} Z_i = 8.2\times 10^{-16} \qquad \text{Var}(\bar{Z}) = 0.914$$



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 $1 \le i \le n$)
    - By Central Limit Theorem, Z ~ N(0, 1), where:

$$Z_n = \frac{\left(\sum_{i=1}^n X_i\right) - n\mu}{\sigma\sqrt{n}} = \frac{\left(\sum_{i=1}^n X_i\right) - nt}{2\sqrt{n}}$$

$$P(-0.5 \le \frac{\sum_{i=1}^n X_i}{n} - t \le 0.5) = P(\frac{-0.5\sqrt{n}}{2} \le \frac{\sqrt{n}}{2}\frac{\left(\sum_{i=1}^n X_i\right) - nt}{n} \le \frac{0.5\sqrt{n}}{2}) = P(\frac{-0.5\sqrt{n}}{2} \le Z_n \le \frac{0.5\sqrt{n}}{2})$$

$$= \Phi(\frac{\sqrt{n}}{4}) - \Phi(\frac{-\sqrt{n}}{4}) = \Phi(\frac{\sqrt{n}}{4}) - (1 - \Phi(\frac{\sqrt{n}}{4})) = 2\Phi(\frac{\sqrt{n}}{4}) - 1 \approx 0.95 \ \Rightarrow\ \Phi(\frac{\sqrt{n^*}}{4}) = 0.975$$

    - Solve for n*: $\frac{\sqrt{n^*}}{4} = 1.96 \ \Rightarrow\ n^* = \lceil (7.84)^2 \rceil = 62$

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 $1 \le i \le n$)
  - What would Chebyshev say? $P(|X_S - \mu_S| \ge k) \le \frac{\sigma_S^2}{k^2}$

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

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

    - Thanks for playing Pafnuty...

Crashing Your Web Site

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

$$P(X \ge 120) = P(X \ge 119.5) = P(\frac{X - 100}{\sqrt{100}} \ge \frac{119.5 - 100}{\sqrt{100}}) = 1 - \Phi(1.95) \approx 0.0256$$

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

$$P(X \ge 120) = P(X \ge E[X] + a) \le \frac{\sigma^2}{\sigma^2 + a^2} = \frac{100}{100 + 20^2} = 0.2$$

It's play time!

Sum of Dice

- You will roll 10 6-sided dice ($X_1, X_2, \ldots, X_{10}$)
  - X = 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):

$$E[X] = 10E[X_i] = 10(3.5) = 35 \qquad \text{Var}(X) = 10\,\text{Var}(X_i) = 10\frac{35}{12} = \frac{350}{12}$$

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

$$\approx 1 - (2\Phi(1.76) - 1) \approx 2(1 - 0.9608) = 0.0784$$

  - If only Chebyshev were right...

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