Announcements

PS5 due today
- Pain poll

PS6 out today
- Due next Monday 8/13 (1:30pm) (will not be accepted after Wed 8/15)
- Programming part: Java, C, C++, or Python

Lecture Notes published for all material
Summary from last time

\[ P(X \geq a) \leq \frac{E[X]}{a} \]
for all \( a > 0 \) if \( X \geq 0 \)

(Markov’s inequality)

\[ P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2} \]
for all \( k > 0 \)

if \( E[X] = \mu, \text{Var}(X) = \sigma^2 \)

(Chebyshev’s inequality)

\[ P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2} \]
for all \( a > 0 \)

if \( E[X] = 0, \text{Var}(X) = \sigma^2 \)

(One-sided Chebyshev’s inequality)

\[ E[f(X)] \geq f(E[X]) \]
if \( f'''(x) \geq 0 \) for all \( x \)

(Jensen’s inequality)
Bounds on sample mean

Consider a sample of \( n \) i.i.d random variables \( X_1, X_2, \ldots, X_n \), where \( X_k \) has distribution \( F \) with \( E[X_k] = \mu \) and \( \text{Var}(X_k) = \sigma^2 \).

Sample mean

\[
\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k
\]

\[
E[\bar{X}] = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}
\]
Summary from last time

Weak Law of Large Numbers
\[
\lim_{n \to \infty} P \left( |\bar{X} - \mu| \geq \varepsilon \right) = 0
\]
→ “probability goes to zero in the limit”
→ Will still have a deviation in the limit (at most \(\varepsilon\))
→ We have an infinite number of non-zero probabilities

Strong Law of Large Numbers
\[
P \left( \lim_{n \to \infty} \bar{X} = \mu \right) = 1
\]
→ Implies WLLN
→ After some number \(n\), \(P (|\bar{X} - \mu| \geq \varepsilon) = 0\)
→ We have a finite number of non-zero probabilities
Where are we going with all of this?

As engineers, we want to:

1. Model things that can be random.
2. Find average values of situations that let us make good decisions.

We perform multiple experiments to help us with engineering. Counts and averages are both sums of RVs.

Weeks 4 and 5:
Learning how to model multiple RVs

This week (week 6):
• Finding expectation of complex situations
• Defining sampling and using existing data
• Mathematical bounds w.r.t modeling the average

Week 7: bringing it all together
• Sampling + average = Central Limit Theorem
• Samples + modeling = finding the best model parameters given data
Goals for today

Central Limit Theorem!
- Confidence intervals, re-defined
- Estimates for sums of IID RVs

Introduction to Parameter estimation
Central Limit Theorem

If $E[X_i] = \mu, \text{Var}(X_i) = \sigma^2$,

Sample means of IID RVs are normally distributed.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N \left( \mu, \frac{\sigma^2}{n} \right)$$

Sums of IID RVs are normally distributed.

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$
CLT explains a lot

$X \sim \text{Bin}(n = 200, p = 0.6)$

Normal approximation to Binomial when $np(1-p) > 10$
A short history of the CLT

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: Alexandr Lyapunov provides precise definition and rigorous proof of CLT

2012: Drake releases “Started from the Bottom”
  - As of July 4th, he has a total of 190 songs
  - Mean quality of subsamples of songs is normally distributed (thanks to the Central Limit Theorem)
Implications of CLT

Anything that is a sum/average of independent random variables is normal...

...meaning in real life, many things are normally distributed:

- Exam scores: sum of individual problems
- Movie ratings: averages of independent viewer scores
- Polling:
  - Ask 100 people if they will vote for candidate X
    - $p_1 = \# \text{“yes”}/100$
    - Sum of Bernoulli RVs (each person independently says “yes” w.p. $p$)
  - Repeat this process with different groups to get $p_1, ..., p_n$ (different sample statistics)
  - Normal distribution over sample means $p_k$
  - Confidence interval: “How likely is it that an estimate for true $p$ is close?”
No more convolution when $n \to \infty!$

For independent discrete random variables $X$ and $Y$, and $Z = X + Y$,

$$p_Z(z) = P(X + Y = z) = \sum_x p_X(x)p_Y(z - x) = \sum_y p_X(z - y)p_Y(y)$$

$p_Z$ is defined as the convolution of $p_X$ and $p_Y$.

Sums of IID RVs are normally distributed.

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

(caveat: must be independent and identically distributed)
CLT vs. Water speed stat

Water Pokemon speed statistic, $X_i$ :

- $E[X_i] = 65$, $\text{Var}(X_i) = 510$
- 126 Water Pokemon
- Let $Y = \text{average of 50 Water Pokemon}$.

What is the distribution of $Y$?

Solution:

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

$$E[Y] = E[X_i] = 65 \quad \text{and} \quad Y \sim N(65, 10.2)$$

$$\text{Var}(Y) = \frac{\text{Var}(X_i)}{n} = \frac{510}{50} = 10.2$$
Confidence interval, fully defined

Consider a sample of IID RVs $X_1, X_2, \ldots, X_n$, where $n$ is large,

$$
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2, \quad \text{SE} = \frac{S}{\sqrt{n}}
$$

def For large $n$, a $100(1-\alpha)$% confidence interval of $\mu$ is defined as:

$$
\left( \bar{X} - z_{\alpha/2} \cdot (\text{SE}), \bar{X} + z_{\alpha/2} \cdot (\text{SE}) \right)
\n\bar{X} \pm z_{\alpha/2} \cdot (\text{SE})
$$

where $\Phi(z_{\alpha/2}) = 1 - (\alpha/2)$.

ex: 95% confidence interval:

$$
\bar{X} \pm 1.96(\text{SE})
$$

$\alpha = 0.05, \alpha/2 = 0.025 \quad \Phi(z_{\alpha/2}) = 0.975, z_{\alpha/2} = 1.96$

In other words: 95% of time that we sample, true $\mu$ will be in interval $\bar{X} \pm 1.96(\text{SE})$

NOT: $\mu$ is 95% likely to be in this particular interval.
SE with confidence interval

Idle CPUs are the bane of our existence.

- A large company monitors 225 computers (single CPU) for idle hours in order to estimate the average number of idle hours per CPU.
- Sample had $\overline{X} = 11.6$ hrs, $S^2 = 16.81$ hrs$^2$ (so $S = 4.1$ hrs)

Give the 90% confidence interval for the estimate of $\mu$, mean idle hrs/CPU.

Solution:

\[
\alpha = 0.10, \alpha/2 = 0.05 \quad \Phi(z_{\alpha/2}) = 0.95, z_{\alpha/2} = 1.645
\]

Standard error, $SE = \frac{S}{\sqrt{n}} = \frac{4.1}{\sqrt{225}} = \frac{4.1}{15}$

90% confidence interval: 
\[
\overline{X} \pm z_{\alpha/2}(SE) \rightarrow 11.6 \pm 1.645 \left(\frac{4.1}{15}\right)
\]

(11.15, 12.05)

Interpret: 90% of time that such an interval is computed, the true $\mu$ is it.
Break

Attendance: tinyurl.com/cs109summer2018
Sampling distribution

CLT tells us sampling distribution of $\bar{X}$ is approx. normal when $n$ is large.

“large” $n$: \hspace{1cm} n > 30

(larger is better): \hspace{1cm} n > 100

We can also use CLT to decide values of $n$ for good confidence intervals

http://onlinestatbook.com/stat_sim/sampling_dist/
Estimating clock running time

Want to test runtime of algorithm.

- Run algorithm \( n \) times and measure average time
- Know variance is \( \sigma^2 = 4 \text{ sec}^2 \)
- Want to estimate mean runtime: \( \mu = t \)

How large should \( n \) be so that \( P(\text{average time is within } 0.5 \text{ of } t) = 0.95? \)

Solution:

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N \left( t, \frac{\sigma^2}{n} \right) \\
(\bar{X} - t) \sim N \left( 0, \frac{4}{n} \right) \quad \text{(SD: } \frac{2}{\sqrt{n}} \text{)} \\
Z = \frac{\bar{X} - t}{2/\sqrt{n}} \sim N(0,1)
\]
Estimating clock running time

Want to test runtime of algorithm.
• Run algorithm $n$ times and measure average time
• Know variance is $\sigma^2 = 4$ sec$^2$
• Want to estimate mean runtime: $\mu = t$

How large should $n$ be so that $P(\text{average time is within 0.5 of } t) = 0.95$?

Solution:

$$P(t - 0.5 \leq \bar{X} \leq t + 0.5) = P(-0.5 \leq \bar{X} - t \leq 0.5) = P\left(-\frac{0.5\sqrt{n}}{2} \leq Z \leq \frac{0.5\sqrt{n}}{2}\right) = 0.95$$

$$= \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = \Phi\left(\frac{\sqrt{n}}{4}\right) - \left(1 - \Phi\left(\frac{\sqrt{n}}{4}\right)\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 = 0.95$$

$$\Rightarrow \Phi\left(\frac{\sqrt{n^*}}{4}\right) = 0.975$$

Solve for $n^*$. $\frac{\sqrt{n^*}}{4} = 1.96 \Rightarrow n^* = \left[(7.84)^2\right] = 62$

What say you, Chebyshev?

$$P(\left|\bar{X} - t\right| \leq 0.5) \geq 0.95 \Rightarrow P(\left|\bar{X} - t\right| \geq 0.5) \leq \frac{4/n}{(0.5)^2} \leq .05$$

$$16/n \leq 0.05 \Rightarrow n \geq 320$$

Thanks for playing, Pafnuty!
The power of the central limit theorem

Sample means of IID RVs are normally distributed.

\[ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N \left( \mu, \frac{\sigma^2}{n} \right) \]

- Can estimate confidence interval
- Can give probabilities on sample mean

Sums of IID RVs are normally distributed.

\[ \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) \]

- Can estimate any sum of IID RVs (if \( \mu, \sigma^2 \) known)
Crashing website

Let \( X \) = number of visitors to a website/minute: \( X \sim \text{Poi}(100) \)

- The server crashes if \( \geq 120 \) requests/minute.

What is \( P(\text{crash in next minute}) \)?

**Solution 1**: (exact, using Poisson)

\[
P(X \geq 120) = \sum_{k=120}^{\infty} \frac{e^{-100}(100)^k}{k!} \approx 0.0282
\]

**Solution 2**: (using CLT)

Recall: sum of IID Poissons is Poisson

CLT says sums are normally distributed (as \( n \to \infty \))

Let:

\[
P(100) \sim \sum_{k=1}^{n} \text{Poi}(100/n)
\]

Using CLT on discrete sum \( \to \) continuity correction

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

**Attempted solution 3**: (using one-sided Chebyshev)

\[
P(X \geq 120) = P(X \geq E[X] + a) \leq \frac{\sigma^2}{\sigma^2 + a^2} = \frac{100}{100 + 20^2} = 0.2
\]
Summary: Approximations

**Poisson:**
- Poisson = sum of IID Poissons where $n$ is arbitrarily large

**Binomial:**
- $n$ trials
- $P(\text{success}) = p$
- Independent trials

**Sum of IID RVs**
- with known $E[X]$, $\text{Var}(X)$

**Normal:**
- (by Central Limit Theorem)

- $p$ medium, $np(1-p) > 10$
- Independent trials

**Poisson:**
- $p < 0.05$, $n > 20$ AND/OR slight dependence
Another dice game

You will roll 10 6-sided dice (X₁, X₂, ..., X₁₀)
Let: \( X = \text{total value of all 10 dice} = X₁ + X₂ + ... + X₁₀ \)
Win if: \( X \leq 25 \) or \( X \geq 45 \)
What is \( P(\text{win}) \)?
Another dice game

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

Let: \(X = \text{total value of all 10 dice} = X_1 + X_2 + \ldots + X_{10}\)

Win if: \(X \leq 25\) or \(X \geq 45\)

What is \(P(\text{win})\)?

Solution:

WTF:

\[
P(X \leq 25 \text{ or } X \geq 45) = 1 - P(25 < X < 45)
\]

By CLT, \(X = \sum_{i=1}^{n} X_i\) approx. \(N(n\mu, n\sigma^2)\) where \(E[X_i] = \mu = 3.5, \text{Var}(X_i) = \sigma^2 = \frac{35}{12}\)

\[
X \sim N((10)3.5 = 35, (10)35/12 = 350/12)
\]

\[
1 - P(25 < X < 45) \approx 1 - P(25.5 \leq X \leq 44.5) \quad \text{continuity correction}
\]

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

\[
\approx 1 - (2\Phi(1.76) - 1) \approx 2(1 - 0.9608) = 0.0784
\]
Another dice game

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

Let: \( X = \text{total value of all 10 dice} = X_1 + X_2 + \ldots + X_{10} \)

Win if: \( X \leq 25 \) or \( X \geq 45 \)

What is \( P(\text{win})? \)

Solution:

\[
P(X \leq 25 \text{ or } X \geq 45) = 1 - P(25 < X < 45) = 0.0784
\]
Finding Statistics

What do you know?

Distribution with parameters

Sample of population

E[X] or Var(X)

RVs

Bootstrapping

CLT

What statistic can you find?

Anything your heart desires

Estimates of E[X] or Var(X), confidence interval, p-values

Sum of IID RVs?

Probability bounds
Now...

What do you know?  
Distribution with parameters

RVs

What statistic can you find?  
Anything your heart desires

What if you don’t know the parameters of your distribution?
## Parameter estimation

**def** parametric model – probability distribution that has parameters \( \theta \).

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Model</th>
<th>Parameters ( \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ber((p))</td>
<td>Bernoulli</td>
<td>( \theta = p )</td>
</tr>
<tr>
<td>Poi((\lambda))</td>
<td>Poisson</td>
<td>( \theta = \lambda )</td>
</tr>
<tr>
<td>Multinomial((p_1, p_2, \ldots, p_m))</td>
<td>Multinomial</td>
<td>( \theta = (p_1, p_2, \ldots, p_m) )</td>
</tr>
<tr>
<td>Unif((\alpha, \beta))</td>
<td>Uniform</td>
<td>( \theta = (\alpha, \beta) )</td>
</tr>
<tr>
<td>Normal((\mu, \sigma^2))</td>
<td>Normal</td>
<td>( \theta = (\mu, \sigma^2) )</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note \( \theta \) can be a vector of parameters!
Why do we care?

**def** Parameter estimation

1. We observe data that has a known model with unknown parameter $\theta$.
2. Use data to estimate model parameters as $\hat{\theta}$.

**Note:** estimators $\hat{\theta}$ are RVs estimating parameter $\theta$
(e.g., sample mean: $\hat{\theta} = \bar{X}$ *estimates* population mean $\theta = \mu$)

“Point estimate” of parameter:
Find the best single value for parameter estimate (as opposed to distribution of $\hat{\theta}$)
• Better understand of process producing data
• Future predictions based on model
• Simulation of future processes

**Machine Learning:**
Learn model of system
Do stuff with model
Estimator bias

**def** bias of estimator: \( E[\hat{\theta}] - \theta \)

- “unbiased” estimator: bias = 0

**Example** Sample Mean

Estimator: \( \bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k \)

Bias: \( E[\bar{X}] = \mu \) (unbiased estimator)

**Example 2** Sample Variance

Estimator: \( S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \)

Bias: \( E[S^2] = \sigma^2 \) (unbiased estimator)
**Estimator consistency**

**def** consistent estimator: \[ \lim_{n \to \infty} P \left( |\hat{\theta} - \theta| < \varepsilon \right) = 1 \text{ for } \varepsilon > 0 \]

- As we get more data, the estimate \( \hat{\theta} \) should deviate from the true \( \theta \) by at most a small amount.
- Actually known as “weak” consistency

**Example**  
**Sample Mean**

Estimator: \[ \bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k \]

Estimating: \( \mu \)

By strong law of large numbers (SLLN): \[ P \left( \lim_{n \to \infty} \bar{X} = \mu \right) = 1 \]

Implying weak law of large numbers (WLLN): \[ \lim_{n \to \infty} P \left( |\bar{X} - \mu| \geq \varepsilon \right) = 0 \text{ for } \varepsilon > 0 \]

Equivalently:

\[ \lim_{n \to \infty} P \left( |\bar{X} - \mu| < \varepsilon \right) = 1 \text{ for } \varepsilon > 0 \]

(consistent estimator)
Sample variance consistency

**def** biased estimator: \( E[\hat{\theta}] - \theta \neq 0 \)

**def** consistent estimator: \( \lim_{n \to \infty} P \left( |\hat{\theta} - \theta| < \varepsilon \right) \) for \( \varepsilon > 0 \)

As \( n \to \infty \), estimate \( \hat{\theta} \) should deviate from true \( \theta \) by at most a small amount.

Are the below two estimators for variance \( \sigma^2 \) biased? Consistent?

1. \( S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \) (sample variance)

   **unbiased:** \( E[S^2] = \sigma^2 \)

   **consistent:** \( S^2 = \frac{n}{n-1} \left( \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 \right) \),

   \[ P \left( \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i^2 = E[X^2] \right) = 1, \ P \left( \lim_{n \to \infty} \bar{X}^2 = \mu^2 \right) = 1 \quad \text{(by SLLN, } X_i \text{ i.i.d.)} \]

2. \( Y = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \)

   **biased:** \( E[Y] = E\left[ \frac{n-1}{n} S^2 \right] = \frac{n-1}{n} \sigma^2 \)

   **consistent:** for similar reasons as above #1, where now \( \lim_{n \to \infty} \frac{n-1}{n} = 1 \)