MATH180A: Introduction to Probability

This week:

homework #4 (due Friday, Nov 1, 11:59pm)

Today: ASV 4.3

Next: ASV 4.4

www.math.ucsd.edu/~ynemish/180a
**CLT for the binomial**

**Thm (CLT for binomial).** Let $S_n \sim B(n, p)$, $a < b$.

Then

$$
\lim_{{n \to \infty}} P\left( a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right) = \Phi(b) - \Phi(a)
$$

For the average $\frac{S_n}{n}$

$$
P\left( \frac{a \sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{S_n}{n} - p \leq \frac{b \sqrt{p(1-p)}}{\sqrt{n}} \right) \approx \Phi(b) - \Phi(a)
$$

**Rule of thumb:**

$$
np(1-p) > 10
$$

**Thm (LLN for binomial).** $S_n \sim B(n, p)$, Then for any $\varepsilon > 0$

$$
\lim_{{n \to \infty}} P\left( \left| \frac{S_n}{n} - p \right| < \varepsilon \right) = 1
$$
Confidence intervals. Motivation

Independent trials, success rate $p$ (unknown)

$S_n = \text{number of successes after } n \text{ trials}$

$S_n \sim B(n, p)$. E.g.: (biased?) coin tossed $n$ times

LLN: \[ \frac{S_n}{n} \to p, n \to \infty \text{ (in probability)} \]

If $n$ is big, $\frac{S_n}{n}$ is close to $p$.

Usually, we don't know $p$, but we can get a realization of $S_n$ (flipping coin) for finite $n$.

What can we say about $p$?
Confidence intervals. Set-up

CLT: interval \((-a, a)\)

\[
P\left( -a\sqrt{\frac{p(1-p)}{n}} \leq \hat{p} - p \leq a\sqrt{\frac{p(1-p)}{n}} \right) \approx 2\Phi(a) - 1
\]

Questions:

1) For fixed \(n\), find \(\varepsilon\) s.t.

\[
P(1\hat{p} - p \leq \varepsilon) \geq \gamma
\]

2) For fixed \(\varepsilon\), find \(n\) s.t.

\[
P(1\hat{p} - p \leq \varepsilon) \geq \gamma
\]
Confidence intervals

\[ P( p \in [\hat{p} - \varepsilon, \hat{p} + \varepsilon]) \geq \gamma \]

\( [\hat{p} - \varepsilon, \hat{p} + \varepsilon] \) - \( \gamma \)-confidence interval for \( p \)

\( \hat{p} \) is r.v. \( \Rightarrow \) interval is random.

Take some realization of \( \hat{p} \) (a number), say \( \hat{p}_* \)

\( [\hat{p}_* - \varepsilon, \hat{p}_* + \varepsilon] \) - \( \gamma \)-confidence interval of \( p \)

\( \hat{p}_* \) \( \rightarrow \) estimate of \( p \)
Confidence intervals. Computations

• What with unknown $p$ in formula?

$$P\left( |\hat{p} - p| \leq \frac{a \sqrt{p(1-p)}}{\sqrt{n}} \right) \approx 2\Phi(a) - 1 = \gamma$$

The $\gamma$-confidence interval can be taken as

with

and
Confidence intervals. Example 1 (fixed n)

Flip a coin 10000 times.
Number of Heads is 5370.
Compute a 99%-confidence interval for $p = P(\text{Heads})$
Confidence intervals. Example 2 (fixed accuracy)

Flip a potentially biased coin. How many times should we repeat the experiment to be able to compute a 95%-confidence interval for \( p = P(\text{Heads}) \) of length 0.01?

<table>
<thead>
<tr>
<th>( Z )</th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.5000</td>
<td>0.5040</td>
<td>0.5080</td>
<td>0.5120</td>
<td>0.5160</td>
<td>0.5199</td>
<td>0.5239</td>
<td>0.5279</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5398</td>
<td>0.5438</td>
<td>0.5478</td>
<td>0.5517</td>
<td>0.5557</td>
<td>0.5596</td>
<td>0.5636</td>
<td>0.5675</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5793</td>
<td>0.5832</td>
<td>0.5871</td>
<td>0.5910</td>
<td>0.5948</td>
<td>0.5987</td>
<td>0.6026</td>
<td>0.6064</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6179</td>
<td>0.6217</td>
<td>0.6255</td>
<td>0.6293</td>
<td>0.6331</td>
<td>0.6368</td>
<td>0.6406</td>
<td>0.6443</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6554</td>
<td>0.6591</td>
<td>0.6628</td>
<td>0.6664</td>
<td>0.6700</td>
<td>0.6736</td>
<td>0.6772</td>
<td>0.6808</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6915</td>
<td>0.6950</td>
<td>0.6985</td>
<td>0.7019</td>
<td>0.7054</td>
<td>0.7088</td>
<td>0.7123</td>
<td>0.7157</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7257</td>
<td>0.7291</td>
<td>0.7324</td>
<td>0.7357</td>
<td>0.7389</td>
<td>0.7422</td>
<td>0.7454</td>
<td>0.7486</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7580</td>
<td>0.7611</td>
<td>0.7642</td>
<td>0.7673</td>
<td>0.7704</td>
<td>0.7734</td>
<td>0.7764</td>
<td>0.7794</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7881</td>
<td>0.7910</td>
<td>0.7939</td>
<td>0.7967</td>
<td>0.7995</td>
<td>0.8023</td>
<td>0.8051</td>
<td>0.8078</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8159</td>
<td>0.8186</td>
<td>0.8212</td>
<td>0.8238</td>
<td>0.8264</td>
<td>0.8289</td>
<td>0.8315</td>
<td>0.8340</td>
</tr>
<tr>
<td>1.0</td>
<td>0.8413</td>
<td>0.8438</td>
<td>0.8461</td>
<td>0.8485</td>
<td>0.8508</td>
<td>0.8531</td>
<td>0.8554</td>
<td>0.8577</td>
</tr>
<tr>
<td>1.1</td>
<td>0.8643</td>
<td>0.8665</td>
<td>0.8686</td>
<td>0.8708</td>
<td>0.8729</td>
<td>0.8749</td>
<td>0.8770</td>
<td>0.8790</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8849</td>
<td>0.8869</td>
<td>0.8888</td>
<td>0.8907</td>
<td>0.8925</td>
<td>0.8944</td>
<td>0.8962</td>
<td>0.8980</td>
</tr>
<tr>
<td>1.3</td>
<td>0.9032</td>
<td>0.9049</td>
<td>0.9066</td>
<td>0.9082</td>
<td>0.9099</td>
<td>0.9115</td>
<td>0.9131</td>
<td>0.9147</td>
</tr>
<tr>
<td>1.4</td>
<td>0.9192</td>
<td>0.9207</td>
<td>0.9222</td>
<td>0.9236</td>
<td>0.9251</td>
<td>0.9265</td>
<td>0.9279</td>
<td>0.9292</td>
</tr>
<tr>
<td>1.5</td>
<td>0.9332</td>
<td>0.9345</td>
<td>0.9357</td>
<td>0.9370</td>
<td>0.9382</td>
<td>0.9394</td>
<td>0.9406</td>
<td>0.9418</td>
</tr>
<tr>
<td>1.6</td>
<td>0.9452</td>
<td>0.9463</td>
<td>0.9474</td>
<td>0.9484</td>
<td>0.9495</td>
<td>0.9505</td>
<td>0.9515</td>
<td>0.9525</td>
</tr>
<tr>
<td>1.7</td>
<td>0.9554</td>
<td>0.9564</td>
<td>0.9573</td>
<td>0.9582</td>
<td>0.9591</td>
<td>0.9599</td>
<td>0.9608</td>
<td>0.9616</td>
</tr>
<tr>
<td>1.8</td>
<td>0.9641</td>
<td>0.9649</td>
<td>0.9656</td>
<td>0.9664</td>
<td>0.9671</td>
<td>0.9678</td>
<td>0.9686</td>
<td>0.9693</td>
</tr>
<tr>
<td>1.9</td>
<td>0.9713</td>
<td>0.9719</td>
<td>0.9726</td>
<td>0.9732</td>
<td>0.9738</td>
<td>0.9744</td>
<td>0.9750</td>
<td>0.9756</td>
</tr>
<tr>
<td>2.0</td>
<td>0.9772</td>
<td>0.9778</td>
<td>0.9783</td>
<td>0.9788</td>
<td>0.9793</td>
<td>0.9798</td>
<td>0.9803</td>
<td>0.9808</td>
</tr>
</tbody>
</table>
Confidence intervals. Polling

Without going into details (see Example 4.14)

Remark 4.16

Part of population (unknown $p$) prefers product A/
supports candidate B/...

We interview $n$ individuals and $k$ of them give
positive answer (about product/candidate).

What can we say about $p$?

Fix confidence level $\gamma$ (often $\gamma = 0.95$)

Again, $\hat{p} = \frac{k}{n}$ is the estimate of $p$. 
Confidence interval. Exercise (rock vs rap)

You ask 400 randomly chosen people living in SD if they prefer rock or rap. 230 reply that they prefer rock music. Give a 99% confidence interval for the part of the population that prefers rock.
Random walk
CLT. Proof for binomial

Fact 1. (Stirling's formula) As $n \to \infty$, $n! \sim n^n e^{-n} \sqrt{2\pi n}$

Fact 2. For $|x| < 1$, $\log(1+x) = x - \frac{x^2}{2} + O(x^3)$

Proof of CLT. $P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = P\left(a\sqrt{npq} + np \leq S_n \leq b\sqrt{npq} + np\right)$

$$= \sum_{a\sqrt{npq} + np \leq k \leq np + b\sqrt{npq}} P(S_n = k)$$

$$P(S_n = k) =$$
\[
\left( \frac{n^p}{k} \right)^k \left( \frac{n^q}{n-k} \right)^{n-k} =
\]

Take log:

Take exp:
Law of large numbers (LLN)

Let $X_1, \ldots, X_n, \ldots$ be independent and identically distributed and let $E(X_i) = \mu \in \mathbb{R}$. Then for any $\varepsilon > 0$

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

Let $X_i \sim \text{Ber}(p)$.

Thm (LLN for binomial). $S_n \sim \text{B}(n, p)$. Then for any $\varepsilon > 0$

$$\lim_{n \to \infty} P \left( \left| \frac{S_n}{n} - \mu \right| < \varepsilon \right) = 1$$

Binomial today, general later.
Proof