11: Continuity Correction, Sampling Gaussians

Amrita Kaur
July 15, 2022
Announcements

• Midterm goes out today, due Monday!
• No attendance for today
Normal approximates Binomial!
Website testing

- 100 people are given a new website design.
- $X = \#$ people whose time on site increases
- The design actually has no effect, so $P(\text{time on site increases}) = 0.5$ independently.
- CEO will endorse the new design if $X \geq 65$.

What is $P(\text{CEO endorses change})$? *Give a numerical approximation.*

**Approach 1: Binomial**

Define

$X \sim \text{Bin}(n = 100, p = 0.5)$

Want: $P(X \geq 65)$

Solve

$P(X \geq 65) \approx 0.0018$

**Approach 2: Approximate with Normal**

Define

$Y \sim \mathcal{N}(\mu, \sigma^2)$

$\mu = np = 50$

$\sigma^2 = np(1 - p) = 25$

$\sigma = \sqrt{25} = 5$

Solve

$P(X \geq 65) \approx P(Y \geq 65) = 1 - F_Y(65)$

$= 1 - \Phi\left(\frac{65 - 50}{5}\right) = 1 - \Phi(3) \approx 0.0013$
Website testing (with continuity correction)

In our website testing, $Y \sim \mathcal{N}(50, 25)$ approximates $X \sim \text{Bin}(100, 0.5)$.

$P(X \geq 65)$ Binomial

$\approx P(Y \geq 64.5)$ Normal

$\approx 0.0018$

You must perform a continuity correction when approximating a Binomial RV with a Normal RV.
Continuity correction

If $Y \sim \mathcal{N}(np, np(1 - p))$ approximates $X \sim \text{Bin}(n, p)$, how do we approximate the following probabilities?

<table>
<thead>
<tr>
<th>Discrete (e.g., Binomial) probability question</th>
<th>Continuous (Normal) probability question</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = 6)$</td>
<td>$P(5.5 \leq Y \leq 6.5)$</td>
</tr>
<tr>
<td>$P(X \geq 6)$</td>
<td>$P(Y \geq 5.5)$</td>
</tr>
<tr>
<td>$P(X &gt; 6)$</td>
<td>$P(Y \geq 6.5)$</td>
</tr>
<tr>
<td>$P(X &lt; 6)$</td>
<td>$P(Y \leq 5.5)$</td>
</tr>
<tr>
<td>$P(X \leq 6)$</td>
<td>$P(Y \leq 6.5)$</td>
</tr>
</tbody>
</table>
\(X \sim \text{Bin}(n, p)\)
- \(E[X] = np\)
- \(\text{Var}(X) = np(1 - p)\)

\(Y \sim \text{Poi}(\lambda)\)
- \(\lambda = np\)

\(Y \sim \mathcal{N}(\mu, \sigma^2)\)
- \(\mu = np\)
- \(\sigma^2 = np(1 - p)\)
Who gets to approximate?

Poisson approximation
- $n$ large ($> 20$), $p$ small ($< 0.05$)
- slight dependence okay

Normal approximation
- $n$ large ($> 20$), $p$ mid-ranged ($np(1-p) > 10$)
- independence
Stanford Admissions

Stanford accepts 2480 students.
- Each admitted student matriculates w.p. 0.68 (independent trials)
- Let $X = \# \text{ of students who will attend}$

What is $P(X > 1745)$? *Give a numerical approximation.*

Strategy:
- A. Just Binomial
- B. Poisson
- C. Normal
- D. None/other
Stanford Admissions

Stanford accepts 2480 students.

• Each admitted student matriculates w.p. 0.68 (independent trials)
• Let $X = \#$ of students who will attend

What is $P(X > 1745)$? Give a numerical approximation.

Strategy:

A. Just Binomial
B. Poisson
C. Normal
D. None/other

- Just Binomial: computationally expensive (also not an approximation)
- Poisson: $p = 0.68$, not small enough
- Normal: $\sqrt{np(1 - p)} = 540 > 10$
Stanford Admissions

Stanford accepts 2480 students.

- Each admitted student matriculates w.p. 0.68 (independent trials)
- Let \( X = \) # of students who will attend

What is \( P(X > 1745) \)? *Give a numerical approximation.*

**Define an approximation**

Let \( Y \sim \mathcal{N}(E[X], \text{Var}(X)) \)

\[
E[X] = np = 1686 \\
\text{Var}(X) = np(1 - p) \approx 540 \rightarrow \sigma = 23.3
\]

\[
P(X > 1745) \approx P(Y \geq 1745.5)
\]

**Solve**

\[
P(Y \geq 1745.5) = 1 - F(1745.5)
= 1 - \Phi\left(\frac{1745.5 - 1686}{23.3}\right)
\approx 0.0055
\]

\( \Phi \) is the cumulative distribution function of the standard normal distribution.
Sampling Gaussians – with Code
ELO ratings

Basketball == Stats

What is the probability that the Warriors win?
How do you model zero-sum games?
ELO ratings

Each team has an ELO score $S$, calculated based on its past performance.

- Each game, a team has ability $A \sim \mathcal{N}(S, 200^2)$.
- The team with the higher sampled ability wins.

What is the probability that Warriors win this game?

Want:

$P(\text{Warriors win}) = P(A_W > A_O)$
Next week - Is there a better way?

\[ P(A_W > A_O) \]

This is a probability of an event involving \textit{two continuous} random variables!
We’ll learn how to solve this problem analytically next week.

Big goal for next time: Events involving \textit{two discrete} random variables.
Stay tuned!