Two ways to get the binomial probability distribution

Binomial Formula or Table A-1

\[ P(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \]

-  \( n \) factorial = \( n! = n(n-1)(n-2) \cdots (2)(1) \)
  
  \( \text{ex: } 4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24 \) by definition \( 1! = 1 \), \( 0! = 1 \)

\[ \binom{n}{x} = \binom{n}{n-x} = \frac{n!}{(n-x)!x!} = \text{# ways to pick } x \text{ of } n \text{ objects if order doesn't matter} \]

\( \binom{4}{2} = \frac{4!}{(4-2)!2!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 2 \cdot 1} = 6 \)

\( \binom{n}{x} = \text{# ways to get } x \text{ successes in } n \text{ trials} \)

\( \text{ex: # orderings of H and T to get } x \text{ of } n \text{ H's} \)

Ex: Pop Quiz with 5 multiple-choice questions, each with 5 choices. You guess randomly. Let \( Z \) = # guess correctly.

\( Z \sim \text{Bin} (5, .2) \)

\[ P(Z=5) = \frac{5!}{(5-5)!5!} (.2)^5 (.8)^5 = (.2)^5 = 0.0003 \]

\( P(Z=3) = \frac{5!}{(5-3)!3!} (.2)^3 (.8)^5 = \frac{5!}{2!3!} (.2)^3 (.8)^3 = 0.0512 \)

\[ \binom{5}{3} \cdot 3 \cdot 2 \cdot 1 = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 3 \cdot 2 \cdot 1} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 3 \cdot 2 \cdot 1} = 10 \]

\[ \frac{2}{2} = \frac{10}{2} = .5 \]

Binomial Probability Distribution

<table>
<thead>
<tr>
<th>( Z )</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.328</td>
</tr>
<tr>
<td>1</td>
<td>.410</td>
</tr>
<tr>
<td>2</td>
<td>.205</td>
</tr>
<tr>
<td>3</td>
<td>.0512</td>
</tr>
<tr>
<td>4</td>
<td>.006</td>
</tr>
<tr>
<td>5</td>
<td>.0003</td>
</tr>
</tbody>
</table>

What is the prob. you get at least 2 questions right?

\[ P(Z \geq 2) = P(2) + P(3) + P(4) + P(5) \]

\[ = 1 - P(P < 2) = 1 - P(1) - P(0) \]

\[ = 1 - .410 - .328 = .262 \]

\[ P(Z \geq 4) = P(4) + P(5) = .006 + .0003 = .0063 \]

Prob you get an odd number of questions right?

\[ P(Z \text{ is odd}) = P(1) + P(3) + P(5) = .410 + .0512 + 0 = .461 \]
A probability distribution has a theoretical mean and std. dev., like a population mean and population std. dev.

The theoretical mean is called the \textit{expected value}

\[ \text{Expected value } \mu = E[X] = \sum x P(x) \]

Ex: \( X = \text{die roll} \) \( E[X] = (1)(\frac{1}{6}) + (2)(\frac{1}{6}) + (3)(\frac{1}{6}) + (4)(\frac{1}{6}) + (5)(\frac{1}{6}) + (6)(\frac{1}{6}) \)
\[ = \frac{21}{6} = 3 \frac{1}{2} \]

Ex: Brighton Quiz \( E[Z] = (0)(.328) + (1)(.41) + (2)(.205) + (3)(.0512) + (4)(.0006) + (5)(.0003) \)
\[ = 1 \]

\( Y = \# \text{ Hon 2 coin flips} \) \( E[Y] = (0)(\frac{1}{4}) + (1)(\frac{1}{2}) + (2)(\frac{1}{4}) = 1 \)

Theoretical variance / standard deviation

\[ \sigma^2 = \text{Var} (X) = \sum [(x-\mu)^2 P(x)] \quad \sigma = \sqrt{\sum (x-\mu)^2 P(x)} \]

For a binomial (only for a binomial)

\[ \mu = np \]
\[ \sigma = \sqrt{np(1-p)} = \sqrt{npq} \quad \text{where } q = 1-p \]

\[ \sigma_Y = \sqrt{(2)(\frac{1}{2})(\frac{1}{2})} = \sqrt{\frac{1}{2}} \approx .71 \]

\[ \sigma_Z = \sqrt{(5)(2)(.8)} = .89 \]

By the empirical rule, observations more than 2 std dev away from the mean are "unusual"

Events outside \( \mu \pm 2\sigma \)

Observations with probability less than .05 are "unusual"

If a student gets 4 questions right on a 5 question quiz (5 options)

Do we think they guessed randomly?

\( \mu = np = (5)(\frac{1}{5}) = 1 \)
\[ \mu + 2\sigma = 1 + 2(.89) = 2.78 \]

Since 4 > 2.78 so it is unlikely that this student guessed randomly.
Poisson Distribution - When events occur randomly at a certain rate, then the number of events that happen in an interval of time (or space) has a Poisson distribution. Counts may not have a fixed upper limit.

Ex: typos in a paper
    earthquakes
    chips in a cookie
    pepperonis on a pizza slice

Formula: \( P(x) = \frac{m^x e^{-m}}{x!} \), \( e \approx 2.71828 \)

(optional)

Mean is \( m \) - rate times interval
Variance is \( m \) - variance equals mean for a Poisson
Std dev \( \sqrt{m} \)

Ex: Suppose you normally make 4 typos per page, and you type a 9 page paper. Would it be unusual for you to make only 20 typos in the paper?

For a single page, \( m_x = 4 \)
For 9 pages, \( m_Y = (4)(9) = 36 \)
\( \sqrt{m_Y} = \sqrt{36} = 6 \)

\( m_Y - 2\sigma_Y = 36 - 2(6) = 24 \)

So it would be unusual to make fewer than 24 typos in a 9 page paper.
Normal Distribution

Just as there are different numbers of chips per cookie, the weight of each cookie will be slightly different.

Cookie weights are (approximately) normally distributed

Normal = Gaussian = bell curve

Continuous RV with mean $\mu$ and std dev $\sigma$

Symmetric around $\mu$

Standard normal $y = \frac{1}{\sqrt{\pi}} e^{-\frac{x^2}{2}}$

General form: $y = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Compare weights of Chips Ahoy to animal cookies

Different means

Harder to work with different means and std devs so we standardize, work with a standard normal

Standard normal has $\mu = 0$ and $\sigma = 1$

Allows us to use a standardized table - A-I