Introduction to Probability

Probabilities are expressed in terms of *events*

Examples of Events:

1. Six shows on a roll of a die
2. Jack of Spades is drawn from a deck of cards
3. Rains at 3:00 p.m. today in front of the Reitz Union
4. Have an automobile accident this year
5. Florida beat Florida State
6. Florida beat Tennessee
7. Florida beat Florida State and Tennessee
8. Yield of a randomly drawn citrus tree is greater than 8 boxes
9. Mean yield of 25 randomly drawn citrus trees is greater than 8 boxes

Events are denoted by capital letters, A, B, etc.

Probabilities of events are denoted P(A), or P(event occurs)
Determination of Probabilities

Probabilities are determined from:
1. Relative frequency computation
2. Subjective assessment

\[ P(\text{six on roll of die}) = \frac{1}{6}, \text{ a relative frequency computation} \]

\[ P(\text{Florida beat Florida State}) = 0.3, \text{ a subjective assessment} \]

\[ P(\text{Rain today at 3:00 at JWRU}) = 0.4, \text{ combination of relative frequency and subjective assessment} \]

Compound Events and Probabilities of Compound Events

Compound events are formed from combinations of events.

The \textit{union} of events A and B occurs if \textit{either} A or B occur

\[ \{\text{Florida beat FSU}\} \cup \{\text{Florida beat Tennessee}\} = \{\text{Florida beat \textit{either} FSU or Tennessee}\} \]

The \textit{intersection} of events A and B occurs if \textit{both} A and B occur

\[ \{\text{Florida beat FSU}\} \cap \{\text{Florida beat Tennessee}\} = \{\text{Florida beat \textit{both} FSU and Tennessee}\} \]
Venn Diagrams

Universe

Event A

Event B
Calculating Probabilities of Compound Events

\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

\[ P\{\text{Florida beat either FSU or Tennessee}\} = P\{\text{Florida beat FSU}\} + P\{\text{Florida beat Tennessee}\} - P\{\text{Florida beat both FSU and Tennessee}\} \]

Two events, A and B, are *disjoint* if they are mutually exclusive; i.e., if \( A \cap B = \Phi \).

If A and B are disjoint, then \( P(A \cup B) = P(A) + P(B) \)

Two events, A and B, are *independent* if \( P(A \cap B) = P(A)P(B) \)

Are the events \{Florida beat FSU\} and \{Florida beat Tennessee\} independent?

Are the events \{Drives 4WD truck\} and \{Voted Republican\} independent?

Are the events \{Drives Prius\} and \{Voted Republican\} independent?

Are the events \{Lives in Florida\} and \{Voted Republican\} independent?
Conditional Probability

The conditional probability of A given B is the probability that A occurs, when it is known that B occurs. It is calculated by the formula

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} \].

If A and B are independent, then \( P(A) = P(A|B) \)

**Examples**

\[ P(J) = P(\text{draw J from deck of cards}) = \frac{4}{52} = \frac{1}{13} \]

\[ P(C) = P(\text{draw club from deck of cards}) = \frac{13}{52} = \frac{1}{4} \]

Intersection: \( P(J \cap C) = P(J \text{ and } C) = \frac{1}{52} \)

Union:

\[ P(J \cup C) = P(J \text{ or } C) = P(J) + P(C) - P(J \cap C) = \frac{1}{13} + \frac{1}{4} - \frac{1}{52} = \frac{4 + 13 - 1}{52} = \frac{16}{52} \]

\[ P(J \cap Q) = 0 \quad P(J \cup Q) = \frac{8}{52} = \frac{2}{13} \]

Dependence:

\[ P(J \cap C) = P(J)P(C) \quad "J" \text{ and } "C" \text{ are independent} \]

\[ P(J \cap Q) \neq P(J)P(Q) \quad "J" \text{ and } "Q" \text{ are not independent} \]
Application

\[
P(\text{PC}+) = \frac{50}{10000} = \frac{1}{200} = 0.005
\]

\[
P(\text{PSA}+) = \frac{398}{10000} = 0.0398
\]

\[
P(\text{PC}+ \cap \text{PSA}+) = \frac{48}{10000} = 0.0048
\]

Conditional Probability: \( P(\text{A given B}) = P(\text{A}|\text{B}) = \frac{P(\text{A} \cap \text{B})}{P(\text{B})} \)

\[
P(\text{PSA}+ | \text{PC}+) = \frac{48}{50} = 0.96 = \text{sensitivity}
\]

\[
P(\text{PSA}- | \text{PC}-) = \frac{9600}{9950} = 0.9648 = \text{specificity}
\]

\[
P(\text{PC}+ | \text{PSA}+) = \frac{48}{398} = 0.12 = \text{predictive ability}
\]
Random Variables:

Discrete
- # spots on top face of die \((1, 2, 3, 4, 5, 6)\)
- suit of drawn card \((C, D, H, S)\)
- # aphids on leaf \((0, 1, 2, 3, \ldots)\)
- # defects in box of 1000 nails \((0, 1, 2, \ldots, 1000)\)
- # germinating seeds out of 50 \((0, 1, 2, \ldots, 50)\)

Continuous
- heights of people \((0 -- ?)\)
- ph of soil \((0 – 10)\)
- voltage in circuit \((0 -- ?)\)
Binomial Random Variable

Let \( x \) = Number of successes out of \( n \) trials, in which the probability of success on each trial is a number \( \pi \). Then \( x \) has a \textit{binomial} distribution with parameters \( n \) and \( \pi \). This is abbreviated \( x \sim B(n, \pi) \).

Example: Consider flipping a coin, and declare a “success” if a head (H) appears. Then the probability of success is .5. That is, \( \pi = P(S) = .5 \). Suppose the coin is flipped \( n=3 \) times, and \( x = \) number of heads. Then \( x \sim B(3, .5) \). The possible values of \( x \) are \( (0, 1, 2, 3) \). The probability of any event is \( (1/2)^3 = 1/8 \).

<table>
<thead>
<tr>
<th>Events:</th>
<th>HHH</th>
<th>HHT</th>
<th>HTH</th>
<th>THH</th>
<th>HTT</th>
<th>THT</th>
<th>TTH</th>
<th>TTT</th>
</tr>
</thead>
<tbody>
<tr>
<td># Heads</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>P(Event)</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
</tr>
</tbody>
</table>

\[ P(3 \text{ H}) = \frac{1}{8} \quad P(2 \text{ H}) = \frac{3}{8} \quad P(1 \text{ H}) = \frac{3}{8} \quad P(0 \text{ H}) = \frac{1}{8} \]

In general: 

\[ P(k \text{ H}) = \frac{3!}{k!(3-k)!} \left( \frac{1}{2} \right)^3 = \frac{3!}{k!(3-k)!} \left( \frac{1}{8} \right) \]
Binomial Random Variable (con’t)

Example: \( x = \text{number of 1’s in 3 rolls of die} \ (0, 1, 2, 3) \)

<table>
<thead>
<tr>
<th>Events</th>
<th>111</th>
<th>11X</th>
<th>1X1</th>
<th>X11</th>
<th>1XX</th>
<th>X1X</th>
<th>XX1</th>
<th>XXX</th>
</tr>
</thead>
<tbody>
<tr>
<td># 1’s</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>P(Event)</td>
<td>( \frac{1^3}{6^3} )</td>
<td>( \frac{1^2 \cdot 5^1}{6^3} )</td>
<td>( \frac{1^2 \cdot 5^1}{6^3} )</td>
<td>( \frac{1^2 \cdot 5^1}{6^3} )</td>
<td>( \frac{1 \cdot 5^2}{6^3} )</td>
<td>( \frac{1 \cdot 5^2}{6^3} )</td>
<td>( \frac{1 \cdot 5^2}{6^3} )</td>
<td>( \frac{5^3}{6} )</td>
</tr>
</tbody>
</table>

\[
P(3 \text{ 1’s}) = \frac{1 \cdot 1 \cdot 1}{6 \cdot 6 \cdot 6} = \frac{1}{6^3} = \frac{1}{216} = 0.0046
\]

\[
P(k \text{ 1’s}) = \frac{3!}{k!(3-k)!} \left( \frac{1}{6} \right)^k \left( \frac{5}{6} \right)^{3-k}
\]

Binomial Formula

\[\pi = \text{probability of success on single trial}\]

\[
P(y \text{ successes in n trials}) = \frac{n!}{y!(n-y)!} \pi^y (1 - \pi)^{n-y}
\]

Mean of the binomial distribution: \( n\pi \)

Variance of the binomial distribution: \( n\pi(1 - \pi) \)
Normal Distribution and normal random variable:

The notation $y \sim \text{N}(\mu, \sigma^2)$ means “The random variable $y$ is distributed normally with mean $\mu$ and variance $\sigma^2$.

The standard normal distribution has mean $\mu=0$ and variance $\sigma^2=1$. The letter $z$ is reserved to represent the standard normal random variable.

Computer programs and tables are available to obtain probabilities from the normal distribution. For example, you can discover that

- $P(-1 < z < 1) = .68$
- $P(z > 1) = .16$
- $P(-1.96 < z < 1.96) = .95$
- $P(z > 1.96) = .025$
- $P(z > 1.42) = .0778$. 
Using the Normal Distribution

Standardizing a Normal Distribution:

If \( y \sim N(\mu, \sigma^2) \), then \( z = \frac{y - \mu}{\sigma} \sim N(0,1) \).

This result allows us to compute probabilities from any normal distribution using tables or a computer program for the standard normal distribution.

If you wanted to calculate the probability that a random variable \( y \) is greater that 1.42 standard deviations above its mean, you would compute:

\[
P(y > \mu + 1.42\sigma) = P\left( \frac{y - \mu}{\sigma} > 1.42 \right) = P(z > 1.42) = .0778
\]

As a more specific application, suppose you believe the egg weights to be normally distributed with mean 65.4 and standard deviation 5.17. You would calculate the probability that a randomly drawn egg is greater than 70 as:

\[
P(y > 72) = P((y - 65.4) / 5.2) > (72 - 65.4) / 5.2 = P(z > 1.27) = .1
\]
Using the normal distribution—an application

Egg weights are normally distributed with mean $\mu = 65$ (g) and standard deviation $\sigma = 5.0$.

1. What is the probability one randomly drawn egg will exceed:
   a. 65   b. 66   c. 70   d. 75

   Let $y =$ egg weight. Then

   a. $P(y > 65) = P\left(\frac{y - 65}{5} > \frac{65 - 65}{5}\right) = P(z > 0) = \frac{1}{2} = .5$
   b. $P(y > 66) = P\left(\frac{y - 65}{5} > \frac{66 - 65}{5}\right) = P(z > .2) = .4207$
   c. $P(y > 70) = P\left(\frac{y - 65}{5} > \frac{70 - 65}{5}\right) = P(z > 1) = .1587$
   d. $P(y > 75) = P\left(\frac{y - 65}{5} > \frac{75 - 65}{5}\right) = P(z > 2) = .0228$

2. What is the probability one egg is between 66 and 70 g?
   $P(66 < y < 70) = P(y > 66) - P(y > 70) = .4207 - .1587 = .262$
Normal Approximation to the Binomial

You can use the normal distribution to approximate binomial probabilities. This often simplifies a computation. For example, suppose you are shooting free-throws in basketball. You know that you make 75% of your shots; that is, the probability of making any one shot is .75. You have entered a contest that awards a prize if you make at least 18 out of 20 shots. What is the probability that you will win a prize?

You need to calculate $P(y \geq 18)$, where $y$ is the number of shots you make out of 20. The exact probability is given by the binomial formula with $\pi = .75$ and $n = 20$:

$$P(y \geq 18) = P(y = 18) + P(y = 19) + P(y = 20)$$

$$= \frac{20!}{18!2!}.75^{18}.25^2$$

$$+ \frac{20!}{19!1!}.75^{19}.25^1$$

$$+ \frac{20!}{20!0!}.75^{20}.25^0$$

$$= .0069 + .0211 + .0032$$

$$= .0912$$
Normal Approximation to the Binomial

The calculation on the previous page would be tedious by hand, but many computer programs are available that can readily do it. However, even good computer programs may fail for computations involving extremely large $n$.

The normal approximation sets $\mu = n\pi$ and $\sigma^2 = n\pi(1-\pi)$, and assumes $y \sim \text{N}(\mu, \sigma^2)$ to evaluate the probability. The approximation is improved by using a continuity correction, which means you compute $P(y \geq 18-.5) = P(y \geq 17.5)$.

The normal approximation is then computed as:

$\mu = n\pi = 20(.75) = 15$
$\sigma^2 = n\pi(1-\pi) = 20(.75)(.25) = 3.75$
$\sigma = 3.75^{1/2} = 1.94$

$$P(y \geq 17.5) = P((y - 15)/1.94 \geq (17.5 - 15)/1.94)$$

$$= P(z \geq 1.29) = 1 - .901 = .099.$$

This is a reasonable approximation to the exact binomial probability of $P(y \geq 18) = .091$. 
