Probability

- Probability is important to statistics because:
  - study results can be influenced by variation
  - it provides theoretical groundwork for statistical inference
- $0 \leq P(A) \leq 1$
  - In English please: the probability of event A must be between zero and one.
  - Note: $P(A) = Pr(A)$

Random Sampling

- A simple random sample of n items is a sample in which:
  - every member of the population has an equal chance of being selected.
  - the members of the sample are chosen independently.

Example:

Consider our class as the population under study. If we select a sample of size 5, each possible sample of size 5 must have the same chance of being selected.

- When a sample is chosen randomly it is the process of selection that is random.
- How could we randomly select five members from this class randomly?

Table Method (p. 670 in book):
1. Randomly assign id’s to each member in the population $(1 - n)$
2. Choose a place to start in table (close eyes)
3. Start with the first number (must have the same number of digits as n), this is the first member of the sample.
4. Work left, right, up or down, just stay consistent.
5. Choose the next number (must have the same number of digits as n), this is the second member of the sample.
6. Repeat step 5 until all members are selected. If a number is repeated or not possible move to the next following your algorithm.
Random Sampling

Computer Method:
3. Histogram plot (left) and Raw Data index Plot (Right)

Key Issue

- How representative of the population is the sample likely to be?
  - The sample wont exactly resemble the population, there will be some chance variation. This discrepancy is called "chance error due to sampling".
- Definition: **Sampling bias** is non-randomness that refers to some members having a tendency to be selected more readily than others.
  - When the sample is biased the statistics turn out to be poor estimates.

Key Issue

**Example**: Suppose a weight loss clinic is interested in studying the effects of a new diet proposed by one of its researchers. It decides to advertise in the LA Times for participants to come be part of the study.

**Example**: Suppose a lake is to be studied for toxic emissions from a nearby power plant. The samples that were obtained came from the portion of the lake that was the closest possible location to the plant.

Let's Make a Deal Paradox – aka, Monty Hall 3-door problem

- This paradox is related to a popular television show in the 1970's. In the show, a contestant was given a choice of three doors/cards of which one contained a prize (diamond). The other two doors contained gag gifts like a chicken or a donkey (clubs).

Let's Make a Deal Paradox.

- After the contestant chose an initial door, the host of the show then revealed an empty door among the two unchosen doors, and asks the contestant if he or she would like to switch to the other unchosen door. The question is should the contestant switch. Do the odds of winning increase by switching to the remaining door?

- The intuition of most people tells them that each of the doors, the chosen door and the unchosen door, are equally likely to contain the prize so that there is a 50-50 chance of winning with either selection? This, however, is **not the case**.

- The probability of winning by using the switching technique is 2/3, while the odds of winning by not switching is 1/3. The easiest way to explain this is as follows:
Let's Make a Deal Paradox.

- The probability of picking the wrong door in the initial stage of the game is 2/3.
- If the contestant picks the wrong door initially, the host must reveal the remaining empty door in the second stage of the game. Thus, if the contestant switches after picking the wrong door initially, the contestant will win the prize.
- The probability of winning by switching then reduces to the probability of picking the wrong door in the initial stage which is clearly 2/3.

Demos:
- File://C:/Ivo.dir/UCLA_Classes/Applets.dir/SOCR/Prototype1.1/classes/TestExperiment.html
- C:/Ivo.dir/UCLA_Classes/Applets.dir/StatGames.exe

Definitions …

- The law of averages about the behavior of coin tosses—the relative proportion (relative frequency) of heads-to-tails in a coin toss experiment becomes more and more stable as the number of tosses increases. The law of averages applies to relative frequencies not absolute counts of #H and #T.
- Two widely held misconceptions about what the law of averages about coin tosses:
  - Differences between the actual numbers of heads & tails becomes more and more variable with increase of the number of tosses—a seq. of 10 heads doesn’t increase the chance of a tail on the next trial.
  - Coin toss results are fair, but behavior is still unpredictable.

Coin Toss Models

- Is the coin tossing model adequate for describing the sex order of children in families?
  - This is a rough model which is not exact. In most countries rates of B/G is different; form 48% … 52%, usually. Birth rates of boys in some places are higher than girls, however, female population seems to be about 51%.
  - Independence, if a second child is born the chance it has the same gender (as the first child) is slightly bigger.

Data from a “random” draw

- 366 cylinders (for each day in the year) for the US Vietnam war draft. The N-th drawn number, correspond to one B-day, indicating order of drafting.
- So, people born later in the year tend to have lower lottery numbers and a bigger chance of actually being drafted.

Types of Probability

- Probability models have two essential components (sample space, the space of all possible outcomes from an experiment; and a list of probabilities for each event in the sample space). Where do the outcomes and the probabilities come from?
- Probabilities from models – say mathematical/physical description of the sample space and the chance of each event. Construct a fair die tossing game.
- Probabilities from data – data observations determine our probability distribution. Say we toss a coin 100 times and the observed Head/Tail counts are used as probabilities.
- Subjective Probabilities – combining data and psychological factors to design a reasonable probability table (e.g., gambling, stock market).
Sample Spaces and Probabilities

- When the relative frequency of an event in the past is used to estimate the probability that it will occur in the future, what assumption is being made?
  - The underlying process is stable over time;
  - Our relative frequencies must be taken from large numbers for us to have confidence in them as probabilities.
- All statisticians agree about how probabilities are to be combined and manipulated (in math terms), however, not all agree what probabilities should be associated with for a particular real-world event.
- When a weather forecaster says that there is a 70% chance of rain tomorrow, what do you think this statement means? (Based on our past knowledge, according to the barometric pressure, temperature, etc. of the conditions we expect tomorrow, 70% of the time it did rain under such conditions.)

The complement of an event

- The complement of an event $A$, denoted $\overline{A}$, occurs if and only if $A$ does not occur.

$\text{S}$ $A$ $\overline{A}$ $A$ $\overline{A}$

(a) Sample space containing event $A$  (b) Event $A$ shaded  (c) $\overline{A}$ shaded

Figure 4.4.1 An event $A$ in the sample space $S$.

Combining events – all statisticians agree on

- “$A$ or $B$” contains all outcomes in $A$ or $B$ (or both).
- “$A$ and $B$” contains all outcomes which are in both $A$ and $B$.

$A$ $B$ $A$ $B$ $A$ $B$ $A$ $B$

(a) Events $A$ and $B$  (b) “$A$ or $B$” shaded  (c) “$A$ and $B$” shaded  (d) Mutually exclusive events

Mutually exclusive events cannot occur at the same time.

Probability distributions

- Probabilities always lie between 0 and 1 and they sum up to 1 (across all simple events).
- $pr(A)$ can be obtained by adding up the probabilities of all the outcomes in $A$.

$$pr(A) = \sum \text{pr}(E)$$

Job losses in the US

<table>
<thead>
<tr>
<th>Reason for Job Loss</th>
<th>Workplace moved/closed</th>
<th>Slack work</th>
<th>Position abolished</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1,703</td>
<td>1,196</td>
<td>548</td>
<td>3,447</td>
</tr>
<tr>
<td>Female</td>
<td>1,210</td>
<td>564</td>
<td>363</td>
<td>2,137</td>
</tr>
<tr>
<td>Total</td>
<td>2,913</td>
<td>1,760</td>
<td>911</td>
<td>5,584</td>
</tr>
</tbody>
</table>

Table 4.4.1 Job Losses in the US (in thousands) for 1987 to 1991
### Table 4.4.1: Proportion of Job Losses from Table 4.4.1

<table>
<thead>
<tr>
<th>Workplace Position</th>
<th>Move/closed</th>
<th>Slack Work</th>
<th>Abolished</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1,703</td>
<td>1,196</td>
<td>548</td>
<td>3,447</td>
</tr>
<tr>
<td>Female</td>
<td>1,207</td>
<td>564</td>
<td>363</td>
<td>2,134</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2,913</td>
<td>1,760</td>
<td>911</td>
<td>5,584</td>
</tr>
</tbody>
</table>

### Properties of Probability Distributions

- A sequence of numbers \(\{p_1, p_2, p_3, \ldots, p_n\}\) is a **probability distribution** for a sample space \(S = \{s_1, s_2, s_3, \ldots, s_n\}\), if \(p(s_k) = p_k\) for each \(1 \leq k \leq n\). The two essential properties of a probability distribution: 
  - \(p_i \geq 0\) 
  - \(\sum p_i = 1\)

- How do we get the probability of an event from the probabilities of outcomes that make up that event?
- If all outcomes are distinct & equally likely, how do we calculate \(p(A)\)? If \(A = \{a_1, a_2, a_3, \ldots, a_n\}\) and \(p(a_1)=p(a_2)=\ldots=p(a_n)=p\), then 
  \[ p(A) = n \times p(a) = np. \]

### Example of Probability Distributions

Tossing a coin twice. **Sample space** \(S = \{HH, HT, TH, TT\}\), for a fair coin each outcome is equally likely, so the probabilities of the 4 possible outcomes should be identical, \(p\). Since, \(p(HH)=p(HT)=p(TH)=p(TT)=p\) and 
\[ p \geq 0; \quad \sum p = 1 \]

- \(p = \frac{1}{4} = 0.25.\)

### Proportion vs. Probability

- How do the concepts of a proportion and a probability differ? A proportion is a partial description of a real population. The probabilities give us the chance of something happening in a random experiment. Sometimes, proportions are **identical to probabilities** (e.g., in a real population under the experiment choose-a-unit-at-random).
- See the **two-way table of counts** (contingency table) on Table 4.4.1, slide 19. E.g., choose-a-person-at-random from the ones laid off, and compute the chance that the person would be a male, laid off due to position-closing. We can apply the same rules for manipulating probabilities to proportions, in the case where these two are identical.

### Rules for Manipulating Probability Distributions

For mutually exclusive events, 
\[ p(A \text{ or } B) = p(A) + p(B) \]
Select an unmarried couple at random – the table proportions give us the probabilities of the events defined in the row/column titles.

**TABLE 4.5.2 Proportions of Unmarried Male-Female Couples Sharing Household in the US, 1991**

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never Married</td>
<td>Divorced</td>
<td>Widowed</td>
<td>Married to other</td>
</tr>
<tr>
<td>Male</td>
<td>Never Married</td>
<td>Divorced</td>
<td>Widowed</td>
<td>Married to other</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.117</td>
<td>.024</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td>.006</td>
<td>.008</td>
<td>.016</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>.021</td>
<td>.022</td>
<td>.003</td>
<td>.016</td>
</tr>
<tr>
<td>Total</td>
<td>.545</td>
<td>.336</td>
<td>.060</td>
<td>.059</td>
</tr>
</tbody>
</table>

Contingency table based on Melanoma histological type and its location

**Melanoma – type of skin cancer – an example of laws of conditional probabilities**

**TABLE 4.6.1: 400 Melanoma Patients by Type and Site**

<table>
<thead>
<tr>
<th>Type</th>
<th>Head and Neck</th>
<th>Trunk</th>
<th>Extremities</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hutchinson's melomotic freckle</td>
<td>22</td>
<td>2</td>
<td>10</td>
<td>34</td>
</tr>
<tr>
<td>Superficial</td>
<td>16</td>
<td>54</td>
<td>115</td>
<td>185</td>
</tr>
<tr>
<td>Nodular</td>
<td>19</td>
<td>33</td>
<td>73</td>
<td>125</td>
</tr>
<tr>
<td>Indeterm inant</td>
<td>11</td>
<td>17</td>
<td>28</td>
<td>56</td>
</tr>
<tr>
<td>Column Totals</td>
<td>68</td>
<td>106</td>
<td>226</td>
<td>400</td>
</tr>
</tbody>
</table>

Contingency table based on Melanoma histological type and its location

**Conditional Probability**

The conditional probability of A occurring given that B occurs is given by

\[ pr(A \mid B) = \frac{pr(A \text{ and } B)}{pr(B)} \]

Suppose we select one out of the 400 patients in the study and we want to find the probability that the cancer is on the extremities given that it is of type nodular: \( P = \frac{73}{125} = \frac{P(c. \text{ on Extremities} \mid \text{Nodular})}{\text{nodular patients}} \)

#nodular patients with cancer on extremities

#nodular patients

**Multiplication rule- what’s the percentage of Israelis that are poor and Arabic?**

\[ pr(A \text{ and } B) = pr(A \mid B)pr(B) = pr(B \mid A)pr(A) \]

0.0728

1.0

0.14

14% of these are Arabic

52% of this 14% are poor

7.28% of Israelis are both poor and Arabic

(0.52 \times 0.14 = 0.0728)

**Figure 4.6.1 Illustration of the multiplication rule.**

A tree diagram for computing conditional probabilities

Suppose we draw 2 balls at random one at a time without replacement from an urn containing 4 black and 3 white balls, otherwise identical. What is the probability that the second ball is black? Sample Spc?

$$P\{\text{2-nd ball is black}\} = \frac{4}{7} \times \frac{3}{6} + \frac{4}{6} \times \frac{3}{7} = \frac{4}{7}.$$
**Classes vs. Evidence Conditioning**

- **Classes**: healthy (NC), cancer
- **Evidence**: positive mammogram (pos), negative mammogram (neg)
- If a woman has a positive mammogram result, what is the probability that she has breast cancer?

\[
P(\text{cancer} | \text{pos}) = \frac{P(\text{pos} | \text{cancer}) \cdot P(\text{cancer})}{P(\text{pos})}
\]

\[
P(\text{pos} | \text{cancer}) = 0.8
\]

\[
P(\text{cancer}) = 0.107
\]

\[
P(\text{cancer} | \text{pos}) = ?
\]

**Proportional usage of oral contraceptives and their rates of failure**

We need to complete the two-way contingency table of proportions

\[
p_0 = \frac{\text{Steril} \cdot \text{ Oral} \cdot \text{ Barrier} \cdot \text{ IUD} \cdot \text{ Sperm. Total}}{1.00}
\]

**Oral contraceptives cont.**

\[
p(\text{Failed and Oral}) = p(\text{Failed} | \text{Oral}) \cdot p(\text{Oral})
\]

**Remarks …**

- In \(p(A | B)\), how should the symbol “ | ” is read given that ...
- How do we interpret the fact that: The event \( A \) always occurs when \( B \) occurs? What can you say about \( p(A | B) \)?
- When drawing a probability tree for a particular problem, how do you know what events to use for the first fan of branches and which events to use for the subsequent branching? (at each branching stage condition on all the info available up to here. E.g., at first branching use all simple events, no prior is available. At 3-rd branching condition of the previous 2 events, etc.)

**TABLE 4.6.4 Table Constructed from the Data in Example 4.6.8**

<table>
<thead>
<tr>
<th>Method</th>
<th>Steril.</th>
<th>Oral</th>
<th>Barrier</th>
<th>IUD</th>
<th>Sperm.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failed</td>
<td>0.38</td>
<td>0.32</td>
<td>0.24</td>
<td>0.03</td>
<td>0.03</td>
<td>1.00</td>
</tr>
<tr>
<td>Total</td>
<td>3.80</td>
<td>3.20</td>
<td>2.40</td>
<td>0.30</td>
<td>0.30</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**TABLE 4.6.5 Number of Individuals Having a Given Mean Absorbance Ratio (MAR) in the ELISA for HIV Antibodies**

<table>
<thead>
<tr>
<th>MAR</th>
<th>Healthy Donor</th>
<th>HIV patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2</td>
<td>202</td>
<td>250</td>
</tr>
<tr>
<td>2.99</td>
<td>73</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td>12+</td>
<td>0</td>
<td>21</td>
</tr>
</tbody>
</table>

Power of a test is: \(1 - p(\text{FNE})\) = \(1 - 0.976\) = 0.024

**Figure 4.6.6** Putting HIV information into the table.
HIV – reconstructing the contingency table

<table>
<thead>
<tr>
<th>Disease status</th>
<th>Test result</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV</td>
<td>0.0098</td>
<td>0.0002</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Not HIV</td>
<td>0.0693</td>
<td>0.9207</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.0791</td>
<td>0.9209</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 4.6.6 Proportions by Disease Status and Test Result

Proportions of HIV infections by country

<table>
<thead>
<tr>
<th>Country</th>
<th>No. AIDS Cases (millions)</th>
<th>pr(HIV)</th>
<th>pr(HIV and Positive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>218.301</td>
<td>0.00864</td>
<td>0.109</td>
</tr>
<tr>
<td>Canada</td>
<td>6.116</td>
<td>0.00229</td>
<td>0.031</td>
</tr>
<tr>
<td>Australia</td>
<td>3.238</td>
<td>0.00193</td>
<td>0.026</td>
</tr>
<tr>
<td>New Zealand</td>
<td>323</td>
<td>0.00095</td>
<td>0.013</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>5.451</td>
<td>0.00095</td>
<td>0.013</td>
</tr>
<tr>
<td>Ireland</td>
<td>142</td>
<td>0.00039</td>
<td>0.005</td>
</tr>
</tbody>
</table>

People vs. Collins

<table>
<thead>
<tr>
<th>Event</th>
<th>Frequency</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellow car</td>
<td>10</td>
<td>Girl with blond hair</td>
</tr>
<tr>
<td>Man with mustache</td>
<td>4</td>
<td>Black man with beard</td>
</tr>
<tr>
<td>Girl with ponytail</td>
<td>10</td>
<td>Interracial couple in car</td>
</tr>
</tbody>
</table>

Statistical independence

- Events A and B are statistically independent if knowing whether B has occurred gives no new information about the chances of A occurring, i.e. if $pr(A | B) = pr(A)$
- Similarly, $P(B | A) = P(B)$, since $P(B|A)=P(B & A)/P(A)=P(A)P(B)/P(A)=P(B)$
- If A and B are statistically independent, then $pr(A and B) = pr(A) \times pr(B)$

Formula summary cont.

- $pr(S) = 1$
- $pr(\overline{A}) = 1 - pr(A)$
- If A and B are mutually exclusive events, then $pr(A or B) = pr(A) + pr(B)$ (here "or" is used in the inclusive sense)
- If $A_1, A_2, \ldots, A_k$ are mutually exclusive events, then $pr(A_1 or A_2 or \ldots or A_k) = pr(A_1) + pr(A_2) + \ldots + pr(A_k)$

Conditional probability

- Definition: $pr(A | B) = \frac{pr(A and B)}{pr(B)}$
- Multiplication formula: $pr(A and B) = pr(B | A)pr(A) = pr(A | B)pr(B)$
### Formula summary cont.

**Multiplication Rule under independence:**
- If \( A \) and \( B \) are independent events, then
  \[ P(A \text{ and } B) = P(A) P(B) \]
- If \( A_1, A_2, \ldots, A_n \) are mutually independent,
  \[ P(A_1 \text{ and } A_2 \text{ and } \ldots \text{ and } A_n) = P(A_1) P(A_2) \ldots P(A_n) \]

### Law of Total Probability

- If \( \{A_1, A_2, \ldots, A_n\} \) are a partition of the sample space (mutually exclusive and \( \bigcup A_i = S \)) then for any event \( B \)
  \[ P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n) \]

### Bayesian Rule

- If \( \{A_1, A_2, \ldots, A_n\} \) are a non-trivial partition of the sample space (mutually exclusive and \( \bigcup A_i = S, P(A_i)>0 \)) then for any non-trivial event \( A \) (\( P(A)>0 \))
  \[ P(A_i | B) = \frac{P(A_i \text{ and } B)}{P(B)} = \frac{P(B | A_i)P(A_i)}{P(B)} = \frac{\sum_{k=1}^{n} P(B | A_k)P(A_k)}{P(B)} \]

### Classes vs. Evidence Conditioning

- **Classes:** healthy (NC), cancer
- **Evidence:** positive mammogram (pos), negative mammogram (neg)

If a woman has a positive mammogram result, what is the probability that she has breast cancer?

\[
P(\text{cancer} | \text{pos}) = \frac{P(\text{cancer} \text{ and } \text{pos})}{P(\text{pos})} = \frac{P(D)P(C|D)}{P(D)}
\]

\[
P(C|P)=\frac{P(P|C)P(C)}{P(P)} = \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.00995} = 0.95
\]

### Bayesian Rule (different data/example!)

<table>
<thead>
<tr>
<th>Test Results</th>
<th>True Disease State</th>
<th>True Disease State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Disease</td>
<td>Disease</td>
</tr>
<tr>
<td>Negative</td>
<td>OK (0.98505)</td>
<td>False Negative II</td>
</tr>
<tr>
<td>Positive</td>
<td>False Positive I (0.00995)</td>
<td>OK (0.00475)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

\[
P(T | \bigcap \neg D) = \frac{P(T | \neg D)P(\neg D)}{P(T | D)P(D) + P(T | \neg D)P(\neg D)} = 0.01 \times 0.995 = 0.00995
\]

\[
\text{Power of Test} = 1 - P(T^C | D) = 0.00025/0.005 = 0.95
\]

\[
\text{Sensitivity: } TP/(TP+FN) = 0.00475/(0.00475+0.00995) = 0.09
\]

\[
\text{Specificity: } TN/(TN+FP) = 0.98505/(0.98505+0.00995) = 0.99
\]
Examples – Birthday Paradox

- The Birthday Paradox: In a random group of N people, what is the change that at least two people have the same birthday?
- E.x., if N=23, P=0.5. Main confusion arises from the fact that in real life we rarely meet people having the same birthday as us, and we meet more than 23 people.
- The reason for such high probability is that any of the 23 people can compare their birthday with any other one, not just you comparing your birthday to anybody else’s.
- There are N-Choose-2 = 20*19/2 ways to select a pair or people. Assume there are 365 days in a year, P(one-particular-pair-same-B-day)=1/365, and
- P(one-particular-pair-failure)=1-1/365 ≈ 0.99726.
- For N=20, 20-Choose-2 = 190. E={No 2 people have the same birthday is the event all 190 pairs fail (have different birthdays)}, then P(E) = P(failure)^190 = 0.99726^190 = 0.59.
- Hence, P(at-least-one-success)=1-0.59=0.41, quite high.
- Note: for N=42, P>0.9 …

The two-color urn model

N balls in an urn, of which there are
M black balls
N – M white balls

Sample n balls and count X = # black balls in sample

We will compute the probability distribution of the R.V. X

The biased-coin tossing model

Perform n tosses and count X = # heads

pr(H) = p

We also want to compute the probability distribution of this R.V. X!

Are the two-color urn and the biased-coin models related? How do we present the models in mathematical terms?

The answer is: Binomial distribution

The distribution of the number of heads in n tosses of a biased coin is called the Binomial distribution.

Binomial(N, p) – the probability distribution of the number of Heads in an N-toss coin experiment, where the probability for Head occurring in each trial is p.

E.g., Binomial(6, 0.7)

For example P(X=0) = P(all 6 tosses are Tails) =

(1 – 0.7)^6 = 0.3^6 = 0.001

Binary random process

The biased-coin tossing model is a physical model for situations which can be characterized as a series of trials where:
- each trial has only two outcomes: success or failure;
- p = P(success) is the same for every trial; and
- trials are independent.

The distribution of X = number of successes (heads) in N such trials is Binomial(N, p)
Sampling from a finite population – Binomial Approximation

If we take a sample of size \( n \)

- from a much larger population (of size \( N \))
- in which a proportion \( p \) have a characteristic of interest, then the distribution of \( X \), the number in the sample with that characteristic,

  is approximately Binomial\((n, p)\).

  (Operating Rule: Approximation is adequate if \( n / N < 0.1 \).)

Example, polling the US population to see what proportion is/has been married.

Binomial Probabilities – the moment we all have been waiting for!

- Suppose \( X \sim \text{Binomial}(n, p) \), then the probability

  \[ P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad 0 \leq x \leq n \]

- Where the binomial coefficients are defined by

  \[ \binom{n}{x} = \frac{n!}{(n-x)! \cdot x!}, \quad n! = 1 \times 2 \times 3 \times \ldots \times (n-1) \times n \]

Expected values

- The game of chance: cost to play: $1.50; Prices: $1, $2, $3; probabilities of winning each price are: 0.6, 0.3, 0.1, respectively.
- Should we play the game? What are our chances of winning/losing?

<table>
<thead>
<tr>
<th>Prize ($)</th>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>( p(x) )</td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

What we would "expect" from 100 games:

- Number of games won:
  - \( 0.6 \times 100 = 60 \)
  - \( 0.3 \times 100 = 30 \)
  - \( 0.1 \times 100 = 10 \)
- Total prize money = Sum; Average prize money = \( \frac{\text{Sum}}{100} \)

\[ \text{Sum} = 1 \times 0.6 + 2 \times 0.3 + 3 \times 0.1 = 1.5 \]

<table>
<thead>
<tr>
<th>Prize ($)</th>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>( p(x) )</td>
<td>0.6</td>
<td>0.3</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

Theoretically Fair Games: price to play EQ the expected return!

Definition of the expected value, in general.

- The expected value:

  \[ E(X) = \sum x P(x) \]

  \[ = \int P(x) \, dx \]

  \( \text{all } x \quad \text{all } X \)

  \[ = \text{Sum of (value times probability of value)} \]

Example

In the at least one of each or at most 3 children example, where \( X = \{\text{number of Girls}\} \) we have:

<table>
<thead>
<tr>
<th>( X )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(x) )</td>
<td>\frac{1}{8}</td>
<td>\frac{5}{8}</td>
<td>\frac{1}{8}</td>
<td>\frac{1}{8}</td>
</tr>
</tbody>
</table>

\[ E(X) = \sum x P(x) \]

\[ = 0 \times \frac{1}{8} + 1 \times \frac{5}{8} + 2 \times \frac{1}{8} + 3 \times \frac{1}{8} \]

\[ = 1.25 \]

The expected value and population mean

\( \mu_e = \mathbb{E}(X) \) is called the mean of the distribution of \( X \).

\( \mu_x = \mathbb{E}(X) \) is usually called the population mean.

\( \mu_e \) is the point where the bar graph of \( P(X = x) \) balances.
Population standard deviation

The population standard deviation is
\[ \text{sd}(X) = \sqrt{E[(X - \mu)^2]} \]

Note that if X is a RV, then \((X-\mu)\) is also a RV, and so is \((X-\mu)^2\). Hence, the expectation, \(E[(X-\mu)^2]\), makes sense.

Population mean & standard deviation

Expected value:
\[ E(X) = \sum xP(X = x) \]

Variance
\[ Var(X) = \sum (x - E(x))^2 P(X = x) \]

Standard Deviation
\[ SD(X) = \sqrt{Var(X)} = \sqrt{\sum (x - E(x))^2 P(X = x)} \]

Example – 3.18

For the Binomial distribution . . . Mean & SD

\[ E(X) = n \cdot p \]
\[ \text{sd}(X) = \sqrt{n \cdot p \cdot (1 - p)} \]

Example – 3.18

Use HistogramChartDemo4: http://socr.ucla.edu/htmls/SOCR_Charts.html

The Normal Distribution

- Recall: in chapter 2 we used histograms to represent frequency distributions.
- We can think of a histogram as an approximation of the true population distribution.
- A smooth curve representing a frequency distribution is called a density curve.
Why is that so?

\[ E(aX + b) = a \ E(X) + b \quad \text{SD}(aX + b) = |a| \ \text{SD}(X) \]

And why do we care?

\[ E(aX + b) = a \ E(X) + b \quad \text{SD}(aX + b) = |a| \ \text{SD}(X) \]

E.g., say the rules for the game of chance we saw before change and the new pay-off is as follows: \{$0, $1.50, $3\}, with probabilities of \{0.6, 0.3, 0.1\}, as before. What is the newly expected return of the game? Remember the old expectation was equal to the entrance fee of $1.50, and the game was fair!

\[ Y = 3(X-1)/2 \]

\{$1, $2, $3\} \rightarrow \{$0, $1.50, $3\},

\[ E(Y) = 3/2 \ E(X) - 3/2 = 3/4 = $0.75 \]

And the game became clearly biased. Note how easy it is to compute \( E(Y) \).

The Normal Distribution

- The normal distribution is described by a unimodal, bell shaped, symmetric density curve

- Area under density curve between \( a \) and \( b \) is equal to the proportion of \( Y \) values between \( a \) and \( b \).
- The area under the whole curve is equal 1.0

A normal density curve can be summarized with the following formula:

\[ f(y) = \frac{1}{\sigma \sqrt{2\pi}} \ \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \]

- Every normal curve uses this formula, what makes them different is what gets plugged in for \( \mu \) and \( \sigma \).
- Each normal curve is centered at \( \mu \) and the width depends on \( \sigma \).
- (small = tall, large = short/wide).

The Standard Scale

- If random variable \( Y \) is normal with mean \( \mu \) and standard deviation \( \sigma \), we write

\[ Y \sim \text{N}(\mu, \sigma^2) \]

- http://www.SOCR.ucla.edu/htmls/SOCR_Distributions.html

Areas under the normal curve

- Because each normal curve is the result of a single formula the areas under the normal curve have been computed and tabulated for ease of use.
- The Standard Scale
  - Any normal curve can be converted into a normal curve with
  - \( \mu = 0 \) and \( \sigma = 1 \). This is called the standard normal.
Areas under the normal curve

- The process of converting normal data to the standard scale is called standardizing.
- To convert $Y$ into $Z$ (a $z$-score) use the following formula:
  \[ Z = \frac{Y - \mu}{\sigma} \]
- What does a $z$-score measure?

Areas under the normal curve

- Table 3 (also in front of book) gives areas under the standard normal curve
- Example: Find the area that corresponds to $z < 2.0$
  - Always good to draw a picture!
- Example: Find the area that corresponds to $z > 2.0$
- Example: Find the area that corresponds to $1.0 < z < 2.0$
- Example: Find the area that corresponds to $z < 2.56$
  - Tables are antiquated → Use tools like SOCR (socr.ucla.edu)

Relationship to the Empirical Rule

- $\mu \pm s \approx 68\%$
- Recall the Empirical Rule $\mu \pm 2s \approx 95\%$
- $\mu \pm 3s \approx 99\%$
- How can we use the standard normal distribution to verify the properties of the empirical rule?
  - The area: $-1 < z < 1 = 0.8413 - 0.1587 = 0.6826$
  - The area: $-2.0 < z < 2.0 = 0.9772 - 0.0228 = 0.9544$
  - The area: $-3.0 < z < 3.0 = 0.9987 - 0.0013 = 0.9974$

Application to Data

**Example**: Suppose that the average systolic blood pressure (SBP) for a Los Angeles freeway commuter follows a normal distribution with mean 130 mmHg and standard deviation 20 mmHg.

Find the percentage of LA freeway commuters that have a SBP less than 100.

- First step: Rewrite with notation!
  \[ Y \sim N(130, 20) \]
Application to Data

- Visually

```
\begin{align*}
Y &\sim \text{Normal}(\mu = 130, \sigma = 10) \\
\phi(z) &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \\
z &< -1.5 = 0.0668
\end{align*}
```

Application to Data

- Try these:
  - What percentage of LA freeway commuters have SBP greater than 155 mmHg?
  - Between 120 and 175?
- Can also be interpreted in terms of probability.
  - What is the probability that a randomly selected freeway commuter will have a SBP less than 100?
  
```
P(Y < 100) = 0.0668
```

Normal approximation to Binomial

- Suppose $Y \sim \text{Binomial}(n, p)$
- Then $Y = Y_1 + Y_2 + \ldots + Y_n$, where $Y_i \sim \text{Bernoulli}(p)$, $E(Y_i) = np$ & $\text{Var}(Y_i) = np(1-p)$
- $E(Y) = np$ & $\text{Var}(Y) = np(1-p)$
- Standardize $Y$:
  - $Z = \frac{Y - np}{\sqrt{np(1-p)}}$
  - By CLT $Z \sim \text{Normal}(0, 1)$
- Normal Approx to Binomial is reasonable when $np \geq 10$ & $n(1-p) > 10$ (p & (1-p) are NOT too small relative to n).

Normal approximation to Binomial – Example

- Roulette wheel investigation:
  - Compute $P(Y \geq 58)$, where $Y \sim \text{Binomial}(100, 0.47)$
    - The proportion of the Binomial(100, 0.47) population having more than 58 reds (successes) out of 100 roulette spins (trials).
  - Since $np = 47 > 10$ & $n(1-p) = 53 > 10$ Normal approx is justified.
  - $Z = \frac{Y - np}{\sqrt{np(1-p)}} = \frac{58 - 100 \cdot 0.47}{\sqrt{100 \cdot 0.47 \cdot 0.53}} = 2.2$
  - $P(Y \geq 58) \approx P(Z \geq 2.2) = 0.0139$
  - True $P(Y \geq 58) = 0.0177$, using SOCR (demo!)
- Binomial approx useful when no access to SOCR available or when N is large!

Assessing Normality

- How can we tell if our data is normally distributed?
  - Several methods for checking normality
    - Mean = Median
    - Empirical Rule
      - Check the percent of data that within 1 sd, 2 sd and 3 sd (should be approximately 68%, 95% and 99.7%).
      - Histogram or dotplot
      - Normal Probability Plot
  - Why do we care if the data is normally distributed?

Normal Probability Plots

- A normal probability plot is a graph that is used to assess normality in a data set.
- When we look at a normal plot we want to see a straight line.
  - This means that the distribution is approximately normal.
  - Sometimes easier to see if a line is straight, than if a histogram is bell shaped.
Normal Probability Plots

- This is how the plot works:
  - We take the data and plot it against normal scores
  - To compute normal scores we take expected values of ordered observations from a sample of size \( n \) that is normally distributed \( N(0,1) \).
  - When we then compare these "normal scores" to the actual \( y \) values on a graph, if the data were normal, we will see our straight line.

Normal Probability Plots

Example: height example from book p.134-135
Suppose we have the height for 11 women.

<table>
<thead>
<tr>
<th>Height (in)</th>
<th>Nscore</th>
</tr>
</thead>
<tbody>
<tr>
<td>61.0</td>
<td>-1.59322</td>
</tr>
<tr>
<td>62.5</td>
<td>-1.06056</td>
</tr>
<tr>
<td>63.0</td>
<td>-0.72791</td>
</tr>
<tr>
<td>64.0</td>
<td>-0.46149</td>
</tr>
<tr>
<td>64.5</td>
<td>-0.22469</td>
</tr>
<tr>
<td>65.0</td>
<td>0.00000</td>
</tr>
<tr>
<td>66.5</td>
<td>0.22469</td>
</tr>
<tr>
<td>67.0</td>
<td>0.46149</td>
</tr>
<tr>
<td>68.0</td>
<td>0.72791</td>
</tr>
<tr>
<td>68.5</td>
<td>1.06056</td>
</tr>
<tr>
<td>70.5</td>
<td>1.59322</td>
</tr>
</tbody>
</table>

Calculated using SOCR, slightly different than formula from text.

Normal Probability Plots

Example (cont'): Normal probability plot

Normal Probability Plots - Simulation

Example: Random Sampling from Normal (0,5): Raw Sample + QQPlot
http://www.socr.ucla.edu/htmls/SOCR_Modeler.html, Data Generation

Normal Probability Plots - Simulation

Example: Random Sampling from Normal (0,5): 100 Raw Sample + Graphs
http://www.socr.ucla.edu/htmls/SOCR_Charts.html, Index & QQ Plots

Normal Probability Plots - Simulation

Example: Random Sampling from Normal (0,5): 100 Raw Sample + Graphs
http://www.socr.ucla.edu/htmls/SOCR_Charts.html, Histogram Plot