I. Expected Value

A. Definition: If the possible outcome of an experiment or random process are real numbers $a_1,\ a_2,\ a_3, ..., a_n$ which occur with probabilities $p_1,\ p_2,\ p_3,\ ...,\ p_n$ then the ***expected value*** of the process is

$$\sum_{i=1}^{n} a_i p_i = a_1 p_1 + a_2 p_2 + ... + a_n p_n$$

B. Example: If $500,000$ people pay \$5.00 each to play a lottery game with a grand prize of \$1,000,000, 10 second prizes of \$1,000 each, $1,000$ third prizes of \$500 each, and $10,000$ fourth prizes of \$10, what is the expected value of a ticket?

1. We assume that the lottery is fair so each of the $500,000$ tickets has the same chance of containing a lottery number that wins a prize.
   Therefore: $p_k = \frac{1}{500{,}000}$ for all $k = 1,\ 2,\ 3,\ ..., 500000$

2. Let the first outcome, $E_1 \equiv$ the ticket wins the grand prize.
   Therefore: a_1 $=$ prize money $-$ cost of ticket
   $=$ \$1,000,000 – \$5.00 $=$ \$999995

3. Let the next 10 outcomes, E_2 through E_{11}, be the events in which the ticket wins a second prize of \$1,000. For each of these tickets a_i $=$ prize money $-$ cost of ticket
   $=$ \$1,000 – \$5.00 $=$ \$995

4. Let the next 1000 outcomes, E_{12} through E_{1011}, be the events in which the ticket wins a third prize of \$500. For each of these tickets a_i $=$ prize money $-$ cost of ticket
   $=$ \$500 – \$5.00 $=$ \$495

5. Let the next 10000 outcomes, E_{1012} through E_{11011}, be the events in which the ticket wins a third prize of \$500. For each of these tickets a_i $=$ prize money $-$ cost of ticket
   $=$ \$10 – \$5.00 $=$ \$5.00

6. The remaining $488,989$ tickets simply lose \$5.00 the events E_{11012} through E_{500000} we have, for each of these tickets:

$$\begin{aligned} a_i &= \text{prize money} - \text{cost of ticket} \\ &= \$0 - \$5.00 = -\$5.00 \end{aligned}$$

7. The expected value of a ticket is, therefore:

$$\sum_{i=1}^{500000} a_i p_i = \sum_{i=1}^{500000} a_i \frac{1}{500000} \quad \text{since each } p_i = \frac{1}{500000}$$

8. So: $\displaystyle\sum_{i=1}^{500000} a_i p_i = \frac{1}{500000}\sum_{i=1}^{500000} a_i$

$$= \frac{1}{500000}\Big(999995 + 10 \times 995 + 1000 \times 495 + 10000 \times 5 + (-5) \times 488989\Big)$$

$$= \frac{1}{500000}\Big(9999950 + 9950 + 495000 + 50000 - 2444945\Big)$$

$$= -1.78$$

9. So, even though the average lottery player may win some prize on occasion, most consistently lose \$1.78 each time they buy a ticket.

II. Recall: Conditional Probability

A. Definition of Conditional Probability:

1. The probability of a given event assuming that another event has already occurred.

2. The probability $p(E)$ of the event E given that the event F has already occurred, $p(E \mid F)$

3. Note that we are not concerned with the probability $p(F)$, only with the probability that the event E occurs given that an event F has occurred.
   a. In general, we solve these problems by using F as the sample space.

   b. Then, for an event E to occur it must belong to $E \cap F$.

   c. Note that this definition includes the case for which $E \subseteq F$.

B. Formal Definition for $p(E \mid F)$

Let E and F be events with $p(F) > 0$ in a sample space S. The **conditional probability** of E given F, denoted by $p(E \mid F)$, is defined to be:

$$p(E \mid F) = p(E \cap F)/p(F)$$

II. Tree Diagram Representation of Conditional Probability

A. Problem Description: An urn contains 5 blue balls and 7 gray balls. Assume that two of these are chosen at random, one after the other, without replacements.

B. Problems:

1. Find the probability that:
   a. Both balls are blue

   b. The first ball is blue and the second is not blue, i.e., gray.

   c. The first ball is gray and the second ball is blue.

   d. Neither ball is blue, i.e., that both are gray.

2. Find the probability that the second ball is blue.

3. Find the probability that at least one of the balls is blue.

4. If this experiment were repeated many times over what would be the expected value of the number of blue balls?

C. Let S denote the sample space of all possible choices of two balls from the urn.

1. Let B_1 denote the probability that the first ball drawn is blue and B_2 the probability that the second ball is blue.

2. Let G_1 denote the probability that the first ball drawn is not blue, i.e., gray and and G_2 the probability that the second ball is gray.

D. Because there are 12 balls of which 5 are blue and 7 are gray

1. The probability that the first ball drawn is blue is:

$$p(B_1) = \frac{5}{12}$$

2. The probability that the first ball drawn is not blue is:

$$p(G_1) = \frac{7}{12}$$

E. If the first ball drawn is blue then the urn contains 4 blue balls and 7 gray balls. Therefore:

1. The probability that the second ball drawn is blue is

$$p(B_2|B_1) = \frac{4}{11}$$

2. The probability that the second ball drawn is gray is

$$p(G_2|B_1) = \frac{7}{11}$$

F. Then: $p(B_2|B_1) = \frac{4}{11}$ and $p(G_2|B_1) = \frac{7}{11}$

G. According to the rule for conditional probability we have:

$$p(B_1 \cap B_2) = p(B_2|B_1) \times p(B_1) = \frac{4}{11} \times \frac{5}{12} = \frac{20}{132}$$

and: $$p(G_2 \cap B_1) = p(G_2|B_1) \times p(B_1) = \frac{7}{11} \times \frac{5}{12} = \frac{35}{132}$$

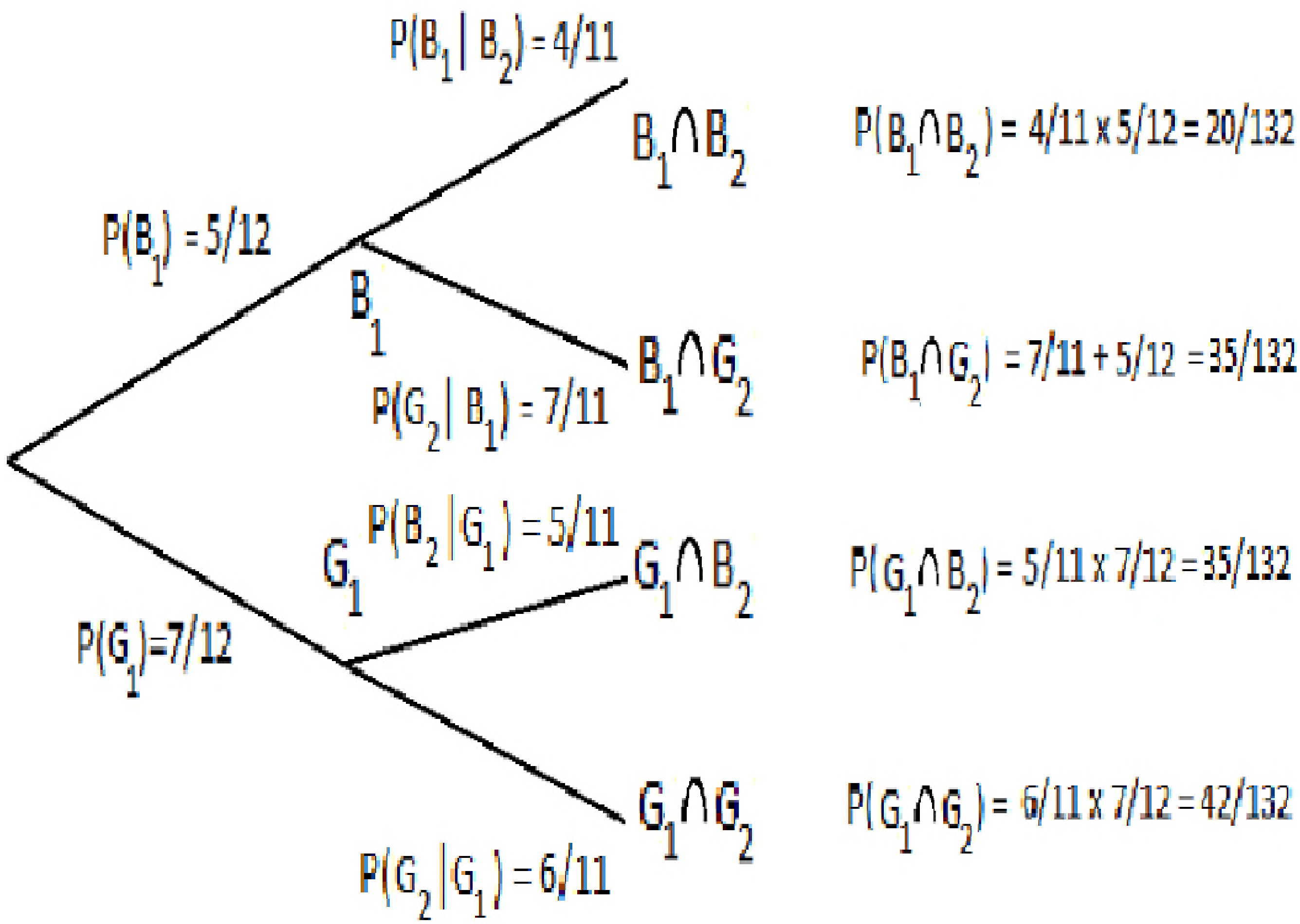
H. If the first ball drawn is not blue we have:

$$p(B_2|G_1) = \frac{5}{11} \quad \text{and} \quad p(G_2|G_1) = \frac{6}{11} \quad \text{which gives:}$$

$$p(G_1 \cap B_2) = p(B_2|G_1) \times p(G_1) = \frac{5}{11} \times \frac{7}{12} = \frac{35}{132}$$

and: $$p(G_1 \cap G_2) = p(G_2|G_1) \times p(B_1) = \frac{6}{11} \times \frac{7}{12} = \frac{42}{132}$$

I. The tree diagram below displays all of the above calculations in place:



J. The event that the second ball is blue can occur in one of two mutually exclusive ways:

1. The first ball is blue and the second is also blue.

2. The first ball is gray and the second is also gray.

3. Therefore: $p(B_2) = p((B_2 \cap B_1) \cup (B_2 \cap G_1))$

$$= p(B_2 \cap B_1) + p(B_2 \cap G_1)$$

$$= \frac{20}{132} + \frac{35}{132} = \frac{55}{132} = \frac{5}{12}$$

4. Note that probability the the second ball is blue is the same as the probability that the first ball is blue.

K. The union of the two events B_1 and B_2, i.e., that at least one of the balls is blue, is given by:

$$p(B_1 \cup B_2) = p(B_1) + p(B_2) - p(B_1 \cap B_2)$$

$$= \frac{5}{12} + \frac{5}{12} - \frac{20}{132} = \frac{90}{132} = \frac{15}{22}$$

L. The event that neither of the balls drawn is blue is the complementt of the event that at least one of the balls drawn is blue.

1. Therefore: $p(G_1 \cap G_2) = 1 - p(B_1 \cup B_2)$

$$= 1 - \frac{15}{22} = \frac{7}{22}$$

2. Equivalently, according to the product rule:

$$p(G_1 \cap G_2) = p(G_1) \times p(G_2)$$

$$= \frac{7}{12} \times \frac{6}{11} = \frac{7}{22}$$

M. To compute the expected value of the number of blue balls that will be drawn we have to consider:

1. The event that one blue ball is drawn which can occur in one of two mutually exclusive ways:

   a. The first ball is blue and the second is not, or:

   $$E_1 = p(B_1 \cap G_2) = p(B_1) \times p(G_2) = \frac{5}{12} \times \frac{7}{11} = \frac{35}{132}$$

   b. The first ball is gray and the second is blue, or:

   $$E_2 = p(G_1 \cap B_2) = p(G_1) \times p(B_2) = \frac{7}{12} \times \frac{5}{11} = \frac{35}{132}$$

   c. Therefore the probability that one blue ball is drawn is, by the summation rule:

   $$p(E_1 \cup E_2) = p(E_1) + p(E_2) = \frac{35}{132} + \frac{35}{132} = \frac{70}{132}$$

2. The event that two blue balls are drawn which can occur in only one way:

   $$p(B_1 \cap B_2) = p(B_1) \times p(B_2) = \frac{5}{12} \times \frac{4}{11} = \frac{20}{132}$$

3. Then the expected value for the number of blue balls is p_{Ex} where:

   $$p_{Ex} = 0 \times P(0 \, \text{blue balls are drawn})$$

   $$+ 1 \times P(1 \text{ blue ball is drawn})$$

   $$+ 2 \times P(2 \text{ blue balls are drawn})$$

   $$= 0 \times \frac{7}{22} + 1 \times \frac{70}{132} + 2 \times \frac{20}{132}$$

   $$= \frac{110}{132} \approx 0.08$$

III. Introduction to Bayes's Theorem: A Problem Example

A. Problem Description:

Suppose that an urn contain 3 blue balls and 4 gray balls. Suppose, further, that a second urn contains 5 blue balls and 3 gray balls. A ball is selected by choosing one of the urns at random and then picking a ball at random from the randomly chosen urn. If the chosen ball is blue what is probability that it came from the first urn?

B. Let:

1. A be the event that the chosen ball is blue.

2. B_1 be the event that the ball came from the first urn.

3. B_2 be the event that the ball came from the second urn.

C. Then: $p(A|B_1) = \frac{3}{7}$ $p(A|B_2) = \frac{5}{8}$

D. Since the urns are equally likely to be chosen we have:

$$p(B_1) = p(B_2) = \frac{1}{2}$$

E. Then: $$p(A \cap B_1) = p(A|B_1) \times\ p(B_1) = \frac{3}{7} \times \frac{1}{2} = \frac{3}{14}$$

and

$$p(A \cap B_2) = p(A|B_2) \times\ p(B_2) = \frac{5}{8} \times \frac{1}{2} = \frac{5}{16}$$

F. But, since A is the disjoint union of the events $(A \cap B_1)$ and $(A \cap B_2)$, we also have that:

$$\begin{aligned} p(A) &= p((A \cap B_1) \cup (A \cap B_2)) \\ &= p(A \cap B_1) + P(A \cap B_2) = \frac{3}{14} + \frac{5}{16} = \frac{59}{112} \end{aligned}$$

G. Finally, using the definition of conditional probability, we have:

$$p(B_1|A) = \frac{p(B_1 \cap A)}{p(A)} = \frac{\frac{3}{14}}{\frac{59}{112}} = \frac{336}{826} \approx 40.7\%$$

H. Therefore: If the chosen ball is blue the probability that it came from the first urn is approximately 40.7%

I. Incidentally: $p(B_2|A) = \frac{p(B_2 \cap A)}{p(A)} = \frac{\frac{5}{16}}{\frac{59}{112}} = 59.3\%$

is the probability that, if the chosen ball is blue, it came from the second urn,

so: $p(B_1|A) + p(B_2|A) = 1$

as we would expect.

V. Bayes's Theorem:

A. Statement: **Suppose that E and F are events from a sample space S such that $p(E) \neq 0$ and $p(F) \neq 0$.**

Then:

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

B. Proof:

1. From the definition of conditional probability we have that:

$$p(F|E)p(E) = p(E|F) \times p(E)$$

or: $p(F|E) = \frac{p(E|F) \times p(E)}{p(E)}$

2. We note that: $E = E \cap S \ = E \cap (F \cup \overline{F})$

$$= (E \cap F) \cup (E \cap \overline{F})$$

3. $(E \cap F)$ and $(E \cap \overline{F})$ must be disjoint.

   Proof (by contradiction):

   a. Assume that $(E \cap F)$ and $(E \cap \overline{F})$ are not disjoint so there exists x such that $x \in E \cap F$ and $x \in E \cap \overline{F}$

   b. Then: $x \in E,$ $x \in F,$ and $x \in \overline{F}$

   c. Therefore: $x \in F \cap \overline{F}$ or $x \in \emptyset$

   d. $x \in \emptyset$ is a contradiction.

   e. Therefore: $(E \cap F)$ and $(E \cap \overline{F})$ must be disjoint

4. As a consequence of $(E \cap F)$ and $(E \cap \overline{F})$ being disjoint we have that:

$$p(E) = p(E \cap F) + p(E \cap \overline{F})$$

5. We have previously shown that:

$$p(E \cap F) = p(E|F)p(F)$$

   and the same proof could be used to show that:

$$p(E \cap \overline{F}) = p(E|\overline{F})p(\overline{F})$$

6. Therefore:

$$\begin{aligned} p(E) &= p(E \cap F) + p(E \cap \overline{F}) \\ &= p(E|F)p(F) + p(E|\overline{F})p(\overline{F}) \end{aligned}$$

7. We substitute this last expression for $p(E)$ into:

$$p(F|E) = \frac{p(E|F) \times p(E)}{p(E)}$$

   to get:

$$p(F|E) = \frac{p(E|F) \times p(E)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

   which is the desired result.

VII. Application of Bayes's Theorem

A. Background:

Most medical tests occasionally produce incorrect results, called false positives and false negative. If a test is desigened to determine if a patient has a certains disease a ***false positive*** indicates that the patient has the disease when, in fact, the patient does not. A ***false negative*** indicates that the patient does not have the disease when the patient does have it.

Large scale health screenings performed for diseases with relatively low incidence have to be designed with several considerations in mind:

1. the per-person cost of the screening.

2. the follow-up costs for further testing of false positives.

3. the possibility that persons who have the disease will develop unwarranted confidences in the state of their health.

B. Problem:

A medical test screens for a disease found in only five people out of $1,000$. The false positive rate is 3% and the false negative rate is 1%. Therefore, 99% of the time a person who has the disease tests positive for it and 97% of the time a person who does not have the disease tests negative for it.

1. What is the probability that a randomly chosen person who tests positive for the disease actually has the disease?

2. What is the probability that a randomly chosen person who tests negative for the disease does not have the disease?

C. Solution:

1. Let F be the event that a randomly chosen person tests positive for the disease.

2. Let E be the event that the chosen person actually has the disease and $\overline{E}$ be the event that the chosen person does not have the disease.

3. We have, then:

   a. $p(F|E) = 0.99$ the probability that a person who has the disease tests positive.

   b. $p(\overline{F}|E) = 0.01$ the probability that a person who has the disease tests negative.

   c. $p(\overline{F}|\overline{E}) = 0.97$ the probability that a person who does not have the disease tests negative.

   d. $p\big(F|\overline{E}\big) = 0.03$ the probability that a person who does not have the disease tests positive.

4. Because five people in 1000 have the disease we have that:

$$p(E) = 0.005 \qquad \text{and} \qquad p(\overline{E}) = 0.995$$

5. By Bayes's Theorem:

   a. $$p(E|F) = \tfrac{p(F|E)\times p(E)}{p(F|E)p(E)+p(F|\overline{E})p(\overline{E})}$$

$$= \tfrac{0.99\times 0.005}{0.99\times 0.005+0.03\times 0.995}$$

$$\approx 0.1422 = 14.2\%$$

   so that the probability that a person who tests positive for the disease actually has the the disease is 14.2%

   b. $$p(\overline{E}|\overline{F}) = \tfrac{p(\overline{F}|\overline{E})\times p(\overline{E})}{p(\overline{F}|E)p(E)+p(\overline{F}|\overline{E})p(\overline{E})}$$

$$= \tfrac{0.97\times 0.995}{0.01\times 0.005+0.97\times 0.995}$$

$$\approx 0.999948 = 99.95\%$$

   so that the probability that a person who tests negative for the disease does not have the disease is 99.95%

6. Comment: The statement that only 14.2% of those who test positive for the disease actually have the disease may be surprising to many of you.

7. Explanation:
   a. Most screening tests are significantly less expensive than a more accurate test.

   b. But, they produce positive results for almost (99.95%) all of those who have the disease.

   c. It is felt to be justifiable to ask those who test positive to undergo the more expensive test.

VII. Generalized Bayes's Theorem:

1. **If a sample space S is a union of mutually disjoint events E_1, E_2, E_3, ..., E_n**

   and

2. **if A is an event in S**

   and

3. **if A and all of the E_i have non-zero probabilities**

   and

4. **if k is an integer with $1 \leq k \leq n$**

   then

5. $$P(E_k|A) = \frac{P(A\,|\,E_k)P(E_k)}{P(A\,|\,E_1)P(E_1)+P(A\,|\,E_2)P(E_2)+\ldots+P(A\,|\,E_n)P(E_n)}$$