STA347 Probability I

Gun Ho Jang

Update on September 9, 2018

Note: This note is prepared for STA347. There might be numerous fault arguments/statements/typos. If you spot one, please contact the instructor or you may look up references which may contain errors too.

Introduction to Probability

The History of Probability

- "The concepts of chance and uncertainty are as old as civilization itself"
- Around 3500 B.C., games of chance played with bone objects.
- Generally believed that the theory of probability was started by Blaise Pascal and Pierre Fermat in 17th century.
- Numerical dice combination probability was studied by Girolamo Cardano and Galileo Galilei in 16th century.

Probability (or statistics) has been employed in many scientific research fields such as

- finance, genetics, human, behavior, marketing, medicine, meteorology, design of computer system
- It probably will rain tomorrow afternoon,
- It is very likely that the plane will arrive late,
- The chances are good that he will be able to join us for dinner this evening.

Interpretations of Probability

Question: Where does the probability come from?

Generally there are two answers. The first view is "frequentism." The following quote is from DS.

> "... the probability that some specific outcome of a process will be obtained can be interpreted to mean the *relative frequency* with which that outcome would be obtained if the process were repeated a *large number* of times *under similar conditions*."

Example 1. If a fair coin were tossed a million times, we would expect head appears around 500,000 times. But if we obtained exactly 500,000 heads, we would be extremely surprised.

- how large a large number meant to be?
- what is the meaning of under similar conditions?
- it must be applied for the same experiment. linear regression may not be applicable

The second view point is "Bayesian".

> Each person has beliefs about random procedure which presumed to be different. It is inferential point of view.
>
> Consider a coin toss problem. If there is no information most people would put equal probability on both head and tail. However, if a person tossed the coin and felt head is more frequently appeared than tail, then the person will consider a higher probability on head. Each person feels different way, hence probability structures vary person by person. In this reason, this concept is called a subjective probability.

Difference of Two School

- Fundamental difference of Frequentism and Bayesian is that,
- in frequentism, samples are coming from one unknown true distribution
- in Bayesian, all unknowns must have probability structure.
- Both are inferential view points.
- There are many other inferential view points
- Most of them accepts one of frequentism or Bayesian aspects.

There is another historical view point which is often called "Classical Interpretation"

> Long ago when a sophisticated probability was not developed yet, probability was discrete and each outcome was treated as equally likely. In a coin toss, head or tail appears with the same probability, 1/2. Each number on a die shows up with the same probability, 1/6.
>
> Obviously many outcomes are not equally likely, e.g., whether is sunny or rainy or cloudy.

Experiments, Events and Sample Spaces

Definition 2. An *experiment* is any process, real or hypothetical, in which the possible outcomes can be identified ahead of time. The collection of all possible outcomes of an experiment is called the *sample space* of the experiment which is often denoted by S. An *event* is a well-defined subset of sample space.

All subsets are events if the sample space is at most countable. Only events are assumed to have probability if the sample space is uncountable. In other words, there are some subsets of sample space which are not events or of which probabilities are not defined. In measure theory, such subsets are called *non-measurable* sets while events are called measurable sets.

Example 2. "Probability of having two heads when a fair coin is tossed 3 times."
Experiment: tossing a fair coin 3 times.
Outcome: TTT, TTH, THT, THH, HTT, HTH, HHT, HHH
Event: two heads $\{THH, HTH, HHT\}$

Exercise 4. Describe experiments, outcomes, events in the following statement. Probability of containing three same cards when five cards are drawn from a shuffled card deck.

Only some subsets of a sample space are events. Suppose $E, E_1, E_2, \ldots$ are events. The followings are also events

(a) E^c

(b) $E_1 \cup E_2 \cup \cdots \cup E_n$

(c) $\bigcup_{i=1}^{\infty} E_i$

Example 3. Consider a study regarding the physical height of UofT students. The event that height is higher than 200cm is the complement of the height is less than or equals to 200cm. The event that height is not between 160cm and 170cm is the union of two events: height less than 160cm and height higher than 170cm. Finally, height less than 170cm is the union of all events that height less than or equal to $(170 - 1/n)$cm for any $n \geq 1$. Thus, all three types of events must be events as described above.

Example 4. Let X be a random number between 0 and 1. $A_r = \{X > r\}$ for $r \in [0,1]$ are events. Then
(a) $B_r = \{X \leq r\} = A_r^c$ are events.
(b) For some $r_1, \dots, r_n \in [0,1]$, the finite union $\bigcup_{i=1}^{n} A_{r_i} = \{X > \min(r_1, \dots, r_n)\} = A_{\min(r_1,\dots,r_n)}$ is also an event.
(c) $C_r = \{X < r\} = \bigcup_{n=1}^{\infty} \{X \leq r - 1/n\} = \bigcup_{n=1}^{\infty} B_{\max(r-1/n,0)}$ are also events.

The Definition of Probability

Consider a set function P on events. Some P satisfies the following axioms.

Axiom 1 $P(E) \geq 0$ for any event E.

Axiom 2 $P(S) = 1$

Axiom 3 For every sequence of disjoint events $E_1, E_2, \dots$

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i).$$

Axioms 1 and 3 are called *non-negativity* and *countable additivity*.

Definition 3. Any function P on a sample space S satisfying Axioms 1-3 is called a *probability*.

Theorem 1. $P(\emptyset) = 0$.

Proof. Let $E_1 = \emptyset = E_2 = E_3 = \cdots$. Then $E_1, E_2, \dots$ are disjoint. Using Axiom 3,

$$P(\emptyset) = P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i) \geq \sum_{i=1}^{2} P(E_i) = 2P(\emptyset)$$

which implies $P(\emptyset) \leq 0$. Axiom 1 implies that $P(\emptyset) \geq 0$. Therefore $P(\emptyset) = 0$. $\square$

Theorem 2 (finite additivity)**.** For any disjoint events $E_1, \dots, E_n$,

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$$

Proof. Define $E_i = \emptyset$ for $i > n$. Then $E_1, E_2, \ldots$ are still disjoint. By Axiom 3,

$$P(\bigcup_{i=1}^{n} E_i) = P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i) = \sum_{i=1}^{n} P(E_i)$$

$\square$

Theorem 3. (a) $P(A^c) = 1 - P(A)$

(b) For $A \subset B$, $P(A) \le P(B)$

(c) $0 \le P(A) \le 1$

(d) $P(A - B) = P(A) - P(A \cap B)$

(e) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(f) [subadditivity, Boole's inequality] For any events $E_1, \ldots, E_n$,

$$P(\bigcup_{i=1}^{n} E_i) \le \sum_{i=1}^{n} P(E_i).$$

Proof. (a) $1 = P(S) = P(A^c \cup A) = P(A^c) + P(A)$ implies $P(A^c) = 1 - P(A)$.

(b) If $A \subset B$, then let $C = B - A$ so that $B = A \cup C$ and $A \cap C = \emptyset$. Then $P(B) = P(A \cup C) = P(A) + P(C) \ge P(A)$.

(c) From $\emptyset \subset A \subset S$, $0 = P(\emptyset) \le P(A) \le P(S) = 1$.

(d) Note that $A = (A - B) \cup (A \cap B)$ and $(A - B) \cap (A \cap B) = \emptyset$. Finite additivity implies $P(A) = P(A - B) + P(A \cap B)$. Thus $P(A - B) = P(A) - P(A \cap B)$.

(e) Note that $A \cup B = A \cup (B - A)$ and $A \cap (B - A) = \emptyset$. Then $P(A \cup B) = P(A) + P(B - A) = P(A) + P(B) - P(A \cap B)$.

(f) For any events $E_1, \ldots, E_n$, define $A_1 = E_1$, $A_i = E_i - \cup_{j=1}^{i-1} E_j$ for $j = 2, \ldots, n$ so that $\cup_{i=1}^{n} A_i = \cup_{i=1}^{n} E_i$ and $A_i \cap A_j = \emptyset$ whenever $i \ne j$. Then $A_i \subset E_i$ implies $P(A_i) \le P(E_i)$. Finally,

$$P(\bigcup_{i=1}^{n} E_i) = P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) \le \sum_{i=1}^{n} P(E_i).$$

$\square$

Example 5. A single card is randomly drawn from a thoroughly shuffled deck of 52 cards. What is the probability that the drawn card will be either a heart or an ace?

Let H and A be the case of a heart and an ace. The question is to compute $P(H \cup A)$ which can be expressed as

$$P(H \cup A) = P(H) + P(A) - P(H \cap A) = 13/52 + 4/52 - 1/52 = 16/52 = 4/13.$$

Theorem 4 (Continuity from below and above)**.** Let P be a probability. (continuity from below). If $A_n \nearrow A$ ($A_1 \subset A_2 \subset \cdots$ and $\cup_n A_n = A$), then $P(A_n) \nearrow P(A)$.
(continuity from above). If $A_n \searrow A$ ($A_1 \supset A_2 \supset \cdots$ and $\cap_n A_n = A$), then $P(A_n) \searrow P(A)$.

Proof. (continuity from below). Let $B_1 = A_1$, $B_n = A_n - A_{n-1}$ for $n \geq 2$. Then $A_n = \cup_{i \leq n} B_i$ and $A = \cup_n B_n$. Hence $P(A_n) = P(A_{n-1}) + P(B_n)$ is non-decreasing and

$$P(A_n) = P(\cup_{i \leq n} B_i) = \sum_{i \leq n} P(B_i) \to \sum_n P(B_n) = P(\cup_n B_n) = P(A).$$

(continuity from above). If $A \subset B$, then $P(B - A) = P(B) - P(A)$. Note $A_1 - A_n \nearrow A_1 - A$. Then $P(A_1) - P(A_n) = P(A_1 - A_n) \nearrow P(A_1 - A) = P(A_1) - P(A)$. Hence $P(A_n) \searrow P(A)$. □

Finite Sample Spaces

Suppose $|S| = n$, that is, $S = \{s_1, \ldots, s_n\}$. Then each member has probability, that is, $p_i = P(\{s_i\})$ such that

$$p_i \geq 0 \text{ and } \sum_{i=1}^{n} p_i = 1.$$

Example 6 (equal probability or classical sense)**.** If each outcome has the same probability, then each element has probability $1/n$, that is, $P(\{s_i\}) = 1/n$ and each event E has probability $P(E) = |E|/n$.

Example 7. When two balanced dice are rolled, what is the probability of sum is 3?
Experiment: rolling two balanced dice
Outcomes: (i, j) with $1 \leq i, j \leq 6$
Probability: $P((i, j)) = 1/36$ for $1 \leq i, j \leq 6$
Event: sum is 3, that is, $E = \{(1, 2), (2, 1)\}$
$P(E) = |E|/36 = 2/36 = 1/18.$

Classical Sense

Example 8. There are three cities A,B,C.

Cities A and B are connected directly in three different routes.

Cities B and C are connected directly 5 different routes.

Cities A and C are not directly connected. There are no other routes passing cities A or B or C.

What is the number of routes from city A to city C through city B?

The choice of routes from A to B doesn't related to the choice of routes from B to C. Hence choose a route from A to B then choose a route from B to C, that is, 3 times 5 is 15.

In general two events are not related number of cases can be multiplied.

Example 9. There are six different books. What is the number of different arrangement of those 6 books?

The first position has 6 possible books. In the second position, remaining 5 books can be placed. And so on. Hence the total number of possible arrangements is $6 \cdot 5 \cdots 2 \cdot 1 = 6! = {}_6P_6 = 6! = 720$.

Example 10. There are n number of balls in a jar, numbered $1, \ldots, n$. When k balls are selected with repetition, what is the probability of having all different numbers?

If $k > n$, it is impossible to draw k different balls. If $k \leq n$, the total number of possible draws are n^k while there are ${}_nP_k$ possible different number draws. Hence the probability is ${}_nP_k/n^k = n!/[(n-k)!n^k]$.

Example 11. When a dice is thrown twice, what is the probability of the sum of two outputs equals 5?

There are $36 = 6 \times 6$ possible combinations. Among them $(1,4), (2,3), (3,2), (4,1)$ are shows sum exactly 5. Hence the probability is $4/36 = 1/9$.

Example 12. Consider an experiment tossing fair coin three times. The sample space is $S = \{HHH, HHT, HTH, HTT, THH$

The probability was taken uniformly or in a classical sense, that is, $P(\{x\}) = 1/|S|$ for all $x \in S$ or $P(A) = |A|/|S|$ for any $A \subset S$. A function $P_2 : 2^S \to [0,1]$ defined as $P(E) = |A \cap \{HHH\}|/2 + |E \cap \{HHH\}^c|/14$ for any $E \subset S$ is also a probability measure.

Exercise 5. Show that P_2 defined in Example 12 is a probability.

Combinatorics

Definition 4. When there are n elements, the number of events pulling k elements out of n elements called a permutation of n elements taken k at a time and denoted by ${}_nP_k$.

Theorem 5. ${}_nP_k = n(n-1)\cdots(n-k+1) = n!/(n-k)!$.

Proof. Out of n elements, the first one can be all n elements. Among remaining $n-1$ elements, the second one can be any of them, that is, $n-1$ possibility. Hence ${}_nP_k = n(n-1) \times \cdots \times (n-k+1) = n!/(n-k)!$. $\square$

Definition 5. The number of combinations of n elements taken k at a time is denoted by ${}_nC_k$ or $\binom{n}{k}$.

Theorem 6. The number of distinct subsets of size k elements taken from a set of n distinct elements is

$${}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

The numbers of permutations and combination have a relationship. From a combination, there are $k!$ permutations which can be generated from the combination, that is, ${}_nP_k = {}_nC_k \times k!$. Hence $\binom{n}{k} = {}_nC_k = {}_nP_k/k! = n!/[k!(n-k)!]$.

Example 13 (Binomial coefficients)**.** When $(x+y)^n$ is expanded there are $(n+1)$ terms, that is, $x^k y^{n-k}$ for $k = 0, \ldots, n$ aside from coefficients. The coefficient is determined by the number of combinations choosing k x-terms among n $(x+y)$-terms. Hence each coefficient is ${}_nC_k$ or $\binom{n}{k}$, that is,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Theorem 7 (Newton Expansion)**.** For $|z| < 1$, the term $(1+z)^r$ can be expanded as

$$(1+z)^r = \sum_{k=0}^{\infty} \binom{r}{k} z^k$$

where the coefficient is defined by

$$\binom{r}{k} = \frac{r(r-1)\cdots(r-k+1)}{k!} = \frac{\Gamma(r+1)}{\Gamma(r-k+1)\Gamma(k+1)}.$$

with $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$.

Example 14. When $x \approx 0$ with $|x| < 1$, $(1+x)^{1/2} = \binom{1/2}{0}x^0 + \binom{1/2}{1}x^1 + \binom{1/2}{2}x^2 + O(|x|^3) = 1 + x/2 - x^2/8 + O(|x|^3)$

Example 15. In a bakery, seven different types of items are sold. A dozen items are packed and sold. How many distinct packages can be sold in the bakery?

The number of packages is the same as the number of choosing $(7-1)$ partitions among $(7-1)$ partition plus 12 indistinguishable items, that is, $\binom{7-1+12}{7-1} = \binom{7-1+12}{12} = 18,564$.

Example 16 ((T) Example 7.5)**.** A certain family has four children. Does the family more likely consist of two children of each sex than of three children of one sex and one of the other?

There are 16 possibilities: BBBB, BBBG, BBGB, BGBB,GBBB, BBGG,GGBB, BGGB, GBBG, BGBG, GBGB, GGGB, GGBG, GBGG, BGGG, and GGGG. Assuming that an unborn child has 50% probability of being a girl and assuming independence of the sex of newborns. Then the probability of having exactly two girls and two boys is $6/16 = 3/8 = 0.375$ while the probability of having three children of one sex and one of the other is $8/16 = 1/2 = 0.5$. Hence it is more likely to have three children in one gender.

Multinomial Coefficients

For any numbers $x_1, \ldots, x_k$ and non-negative integer n,

$$(x_1 + \cdots + x_k)^n = \sum \binom{n}{n_1, \ldots, n_k} x_1^{n_1} \cdots x_k^{n_k}$$

where non-negative integers n_i's satisfying $n_i \geq 0$ and $n_1 + \cdots + n_k = n$.

It is easy to see that

$$\begin{aligned}\binom{n}{n_1, \ldots, n_k} &= \binom{n}{n_1}\binom{n_2 + \cdots + n_k}{n_2}\binom{n_3 + \cdots + n_k}{n_3} \cdots \binom{n_k}{n_k} \\ &= \frac{n!}{n_1! \cdots n_k!}.\end{aligned}$$

Exercise 6. What is the number of terms in the expansion of the theorem?

Example 17 (Birthday Problem)**.** There are k students in a class.
Question: What is the probability of all students' having different birthdays?

There are 365^k birthday combinations. Among them, ${}_{365}P_k$ cases have all different birthdays. Hence the answer is $365!/[(365-k)!365^k]$.

In other words, the probability p there exists at least two students having the same birthday is

$$p = 1 - \frac{365!}{(365-k)!365^k}$$

For some k, this values are

k	5	10	15	20	23	25
p	0.0271	0.1169	0.2529	0.4114	0.5073	0.5687
k	30	40	50	60	80	100
p	0.7063	0.8912	0.9704	0.9941	0.9999	1.0000

Theorem 8 (Stirling's formula)**.** $\lim_{n\to\infty} |\log(n!) - [\frac{1}{2}\log(2\pi) + (n+\frac{1}{2})\log(n) - n]| = 0$.

It is known that

$$\Big|\log(n!) - [\frac{1}{2}\log(2\pi) + (n+\frac{1}{2})\log(n) - n]\Big| \le \frac{1}{12n}$$

Inclusion-Exclusion Formula

Countable additivity implies finite additivity, that is, for any disjoint events $A_1, \dots, A_n$,

$$P(\overset{n}{\underset{i=1}{\cup}} A_i) = \sum_{i=1}^{n} P(A_i).$$

Even though two events A_1, A_2 are not disjoint, there exist two disjoint events of which union is the same to the union of A_1 and A_2. That is, $A_1 \cup A_2 = A_1 \cup (A_2 - A_1)$. Thus

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1 \cup (A_2 - A_1)) = P(A_1) + P(A_2 - A_1) \\ &= P(A_1) + P(A_2) - P(A_1 \cup A_2). \end{aligned}$$

In general the union of finite events can be written as sum of probabilities of intersections.

Theorem 9 (inclusion-exclusion formula)**.** For any n events $A_1, \dots, A_n$,

$$P(\overset{n}{\underset{i=1}{\cup}} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n-1} P(A_1 \cap \cdots \cap A_n).$$

A proof can be obtained using a mathematical induction.

Matching Problem

There are n letters and n envelops for n peoples. Letters are randomly placed into envelops. What is the probability of having at least one correct letter and envelop pair?

Let A_i be the event that the letter i is placed into envelop i. Then the question is $P(\overset{n}{\underset{i=1}{\cup}} A_i)$ which can be computed using inclusion-exclusion formula. First of all $P(A_i) = 1/n$, then $\sum_{i=1}^n P(A_i) = n \times 1/n = 1$. Secondly $P(A_i \cap A_j) = 1/n \times 1/(n-1)$ for $i \neq j$. Hence $\sum_{i<j} P(A_i \cap A_j) = \binom{n}{2} 1/[n(n-1)] = 1/2!$. Inductively $P(A_{i_1} \cap \cdots \cap A_{i_k}) = 1/[n(n-1) \cdots (n-k+1)]$ and $\sum_{i_1 < \cdots < i_k} P(A_{i_1} \cap \cdots \cap A_{i_k}) = \binom{n}{k} 1/[n(n-1) \cdots (n-k+1)] = 1/k!$. Hence $P(\overset{n}{\underset{i=1}{\cup}} A_i) = \sum_{k=1}^n (-1)^{k-1}/k! \to 1 - e^{-1} \approx 0.6321$

Conditional Probability

The Definition of Conditional Probability

Example 18.

- Consider a lottery
- Draw 6 numbers between 1 and 30 without replacement
- Match the numbers without order
- Jack picked 1, 14, 15, 20, 23, 27.
- Compute the probability of winning the lottery?
- Jack saw 15 is being drawn on TV and then lost signal.
- Does the probability of winning the lottery change after learning 15 is one of winning draw?

Definition 6. When $P(B) > 0$, the *conditional probability* of an event A given B is defined by $P(A \,|\, B) = P(A \cap B)/P(B)$.

Example 19. Consider the previous example. Let A be the event drawing 1, 14, 15, 20, 23, 27 and B be the event the winning ticket containing 15.

It is an equal outcome probability problem. There are $\binom{30}{6}$ possible draws. The number of draws containing 15 is $\binom{29}{5}$. Hence

$$P(B) = \binom{29}{5} \Big/ \binom{30}{6} = \frac{29!/(5!24!)}{30!/(6!24!)} = 0.2$$

Considering $A \cap B = A$,

$$P(A \cap B) = P(A) = 1 \Big/ \binom{30}{6} = 6!24!/30! = 1/593,775 \approx 1.6841 \times 10^{-6}$$

Hence the conditional probability becomes

$$P(A \mid B) = \frac{P(A \cap B}{P(B)} = 1/\binom{29}{5} = 1/118,755 \approx 8.4207 \times 10^{-6} > P(A).$$

After watching TV, winning probability is increased.

Actually in this case, it is not difficult to check $P(A \mid B) = 5P(A)$.

Example 20. A fair die is thrown. Given that the number is even, what is the probability that the number is 2?

Clearly $S = \{1, 2, 3, 4, 5, 6\}$. Let $A = \{2\}$ and $B = \{2, 4, 6\}$. The question is $P(A \mid B) = P(A \cap B)/P(B) = P(A)/P(B) = (1/6)/(3/6) = 1/3$.

Under the same condition, what is the probability of the number being 1?
By letting $A = \{1\}$, we get $P(A \mid B) = P(A \cap B)/P(B) = P(\emptyset)/P(2, 4, 6) = (0/6)/(3/6) = 0$. In other words, it is impossible.

Exercise 7. Suppose $P(B) > 0$. Let $Q(A) = P(A \mid B)$. Show that Q satisfies Axioms 1-3 and conclude Q is also a probability.

From the definition of conditional probability, the following theorem is obtained.

Theorem 10. If $P(B) > 0$, then $P(A \cap B) = P(A \mid B)P(B)$.

Example 21. There are r red balls and b blue balls in a box. Two balls are drawn without replacement. What is the probability red and blue balls are drawn in order?

Let A be the event the first drawn ball being red and B the the event the second drawn ball being blue. Then $P(A) = r/(r + b)$ and $P(B \mid A) = b/((r - 1) + b)$. Thus

$$P(A \cap B) = P(A)P(B \mid A) = \frac{r}{r + b} \frac{b}{r + b - 1}$$

Theorem 11. Let $A_1, \ldots, A_n$ be events with $P(A_1 \cap \cdots \cap A_n) > 0$. Then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1, A_2) \cdots P(A_n \mid A_1, \ldots, A_{n-1})$$

Independent Events

Definition 7. Two events A and B are *independent* if and only if $P(A \cap B) = P(A)P(B)$. A collection of events $\{A_i\}_{i\in I}$ are said to be *(mutually) independent* if $P(\cap_{i\in J} A_i) = \prod_{i\in J} P(A_i)$ for any $\emptyset \neq J \subset I$. A collection of events $\{A_i\}_{i\in I}$ are said to be *pair-wise independent* if $P(A_i \cap A_j) = P(A_i)P(A_j)$ for $i \neq j \in I$.

Example 22. Suppose that a balanced die is rolled. Let A be the event that an even number is obtained, and let B be the event that one of the numbers 1, 2, 3, or 4 is obtained.

Note $P(A) = 1/2$ and $P(B) = 2/3$. Also $A \cap B = \{2, 4\}$. Hence $P(A \cap B) = 1/3$. Hence $P(A \cap B) = 1/3 = P(A)P(B)$ and A and B are independent.

Theorem 12. Two events A and B are independent if and only if A and B^c are independent.

Proof. By noting $P(A\cap B^c) = P(A-B) = P(A)-P(A\cap B)$, $P(A\cap B^c)-P(A)P(B^c) = P(A)-P(A\cap B)-P(A)P(B^c) = P(A)[1-P(B^c)]-P(A\cap B) = -[P(A\cap B)-P(A)P(B)]$ Thus $P(A\cap B^c) = P(A)P(B^c)$ if and only if $P(A\cap B) = P(A)P(B)$. Hence A and B are independent if and only if A and B^c are independent. □

Example 23. A fair die is thrown. Consider three events $A = \{\text{it is even}\}$, $B = \{\text{it is less than } 4\}$ and $C = \{\text{it is multiple of } 3\}$. Clearly $A = \{2, 4, 6\}, B = \{1, 2, 3\}, C = \{3, 6\}$. Then $P(A) = P(B) = 1/2$, $P(C) = 1/3$ and $P(A \cap B) = P(\{2\}) = 1/6 \neq 1/4 = P(A)P(B)$. Hence A and B are not independent. While $P(A \cap C) = P(\{6\}) = 1/6 = (1/2)(1/3) = P(A)P(C)$ and $P(B \cap C) = P(\{3\}) = 1/6 = (1/2)(1/3) = P(B)P(C)$ show the independence between A and C; and B and C.

Exercise 8. Suppose there are n events, $E_1, \ldots, E_n$. To check whether or not events are independent, probabilities of many intersections of event must be checked. How many checks are required number to verify independence?

Example 24. Let $S = \{1, 2, 3, 4\}$, $A = \{1, 2\}$, $B = \{1, 3\}$, and $C = \{1, 4\}$. It is easy to see that A, B, C are pair-wise independent. But

$$P(A \cap B \cap C) = P(\{1\}) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C).$$

Hence A, B, C are not (mutually) independent.

Exercise 9. Let $A_1, \ldots, A_n$ be disjoint events. Show that these events are mutually independent if and only if at most one event has positive probability, that is, all events have probability zero or with only one exception.

Definition 8. Two events A and B are *conditionally independent* given C if

$$P(A \cap B \,|\, C) = P(A \,|\, C)P(B \,|\, C).$$

Conditional independence does not imply independence.

Example 25. Consider the sample space $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ with the uniform probability P and events $A = \{1, 2, 3\}, B = \{1, 2, 4, 5\}, C = \{2, 3, 4, 5, 6, 7\}$ so that $P(A) = 1/3, P(B) = 4/9, P(C) = 2/3$, $P(A \cap B) = P(\{1, 2\}) = 2/9, P(A \cap C) = P(\{2, 3\}) = 2/9 = P(A)P(C), P(A \cap B \cap C) = P(\{2\}) = 1/9$, From $P(A \cap C) = P(\{2, 3\}) = 2/9 = P(A)P(C)$, A and C are independent but

$$P(A \cap C \,|\, B) = \frac{P(A \cap B \cap C)}{P(B)} = \frac{1/9}{4/9} = \frac{1}{4} \neq \frac{3}{8} = \frac{2/9}{4/9}\frac{3/9}{4/9} = \frac{P(A \cap B)}{P(B)}\frac{P(C \cap B)}{P(B)} = P(A \,|\, B)P(C \,|\, B)$$

shows A and C are not conditionally independent given B.

Two events A and B are not independent because $P(A \cap B) = 2/9 \neq 4/27 = P(A)P(B)$. While

$$P(A \cap B \,|\, C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{1/9}{6/9} = \frac{1}{6} = \frac{2/9}{6/9}\frac{3/9}{6/9} = \frac{P(A \cap C)}{P(C)}\frac{P(B \cap C)}{P(C)} = P(A \,|\, C)P(B \,|\, C)$$

implies A and B are conditionally independent given B.

Example. Random variables X, Y are conditionally independent with conditional density $1(0 < x < z)/z$ given $Z = z$. And $Z \sim \text{uniform}(0, 1)$. Then margianlly X and Y are not independent by considering

$$P(X, Y < 1/2) = \int_0^1 P(X, Y < 1/2 \,|\, Z = z)\, dz = \int_0^{1/2} 1\, dz + \int_{1/2}^1 \frac{1}{4z^2}\, dz = 1/2 + (-1/4z)|_{1/2}^1 = 1/2 - 1/4 + 1/8 = 3/8.$$
$$P(X < 1/2) = \int_0^1 P(X < 1/2 \,|\, Z = z)\, dz = \int_0^{1/2} 1\, dz + \int_{1/2}^1 \frac{1}{2z}\, dz = 1/2 + \frac{1}{2}\log 2.$$

While $P(X < 1/2, Y < 1/2) = 3/8 < (1 + \log(2))^2/4 = P(X < 1/2)P(Y < 1/2)$ implies the dependence of X and Y.

Bayes' Theorem

Definition 9. A collection of sets $B_1, \ldots, B_k$ is called a *partition* of A if and only if $B_1, \ldots, B_k$ are disjoint and $A = \bigcup_{i=1}^{k} B_i$.

Theorem 13 (Law of total probability)**.** Let events $B_1, \ldots, B_k$ be a partition of S with $P(B_j) > 0$ for all

$j = 1, \ldots, k$. For any event A,

$$P(A) = \sum_{j=1}^{k} P(B_j)P(A \mid B_j)$$

Proof. Note that $\overset{k}{\underset{j=1}{\cup}} B_j = S$ and $A = A \cap S = A \cap \overset{k}{\underset{j=1}{\cup}} B_j = \overset{k}{\underset{j=1}{\cup}} (A \cap B_j)$. Also whenever $i \neq j$ $(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = \emptyset$ or disjoint. Using finite additivity and condition probability,

$$P(A) = P(\overset{k}{\underset{j=1}{\cup}} (A \cap B_j)) = \sum_{j=1}^{k} P(A \cap B_j) = \sum_{j=1}^{k} P(B_j)P(A \mid B_j)$$

$\square$

Example 26. Consider a random number drawing with replacement from 50 numbers $1, \ldots, 50$. The first drawn number is X. Then continue to draw a number Y until $Y \geq X$. What is probability of $Y = 50$?

It is easy to see that $P(Y = 50 \mid X = x) = 1/(51 - x)$. Hence

$$\begin{aligned} P(Y = 50) &= \sum_{x=1}^{50} P(X = x)P(Y = 50 \mid X = x) = \sum_{x=1}^{50} \frac{1}{50} \frac{1}{51 - x} \\ &= \frac{1}{50}\left(1 + \frac{1}{2} + \cdots + \frac{1}{50}\right) \approx 0.08998 \end{aligned}$$

Theorem 14 (Bayes' Theorem)**.** If $0 < P(A), P(B) < 1$, then

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}.$$

Proof. By the definition of conditional probability,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}.$$

The denominator is expanded using the law of total probability. $\square$

Posterior probability can be expressed with respect to prior probabilities.

Example 27. Only two factories manufacture zoggles. 20 per cent of the zoggles from factory A and 5 per cent from factory B are defective. Factory A produces twice as many zoggles as factory B each week. What is the probability that a zoggle, randomly chosen from a weeks production, is satisfactory? Let D be the event that the chosen zoggle is defective, and let A be the event that it was made in factory A. Then

$$P(D^c) = P(D^c \mid A)P(A) + P(D^c \mid A^c)P(A^c) = \frac{4}{5}\cdot\frac{2}{3} + \frac{19}{20}\cdot\frac{1}{3} = 0.85$$

If the chosen zoggle is defective, what is the probability that it came from factory A?

$$P(A \mid D) = \frac{P(A\cap D)}{P(D)} = \frac{P(D \mid A)P(A)}{P(D)} = \frac{\frac{1}{5}\cdot\frac{2}{3}}{1-\frac{51}{60}} = \frac{8}{9} = 0.8889.$$

Example 28 (Simpsons paradox)**.** A doctor has performed clinical trials to determine the relative efficacies of two drugs, with the following results.

	Women		Men	
	Drug I	Drug II	Drug I	Drug II
Success	200	10	19	1000
Failure	1800	190	1	1000

Q: Which drug is the better?

Answer 1. The overall curation rate for Drug I and II are 219/2020 and 1010/2200 respectively. Hence Drug II is better than Drug I.

Answer 2. Amongst women, the curation rates are 1/10 and 1/20 and amongst men 19/20 and 1/2. Hence Drug I is better in both cases.

Note. This phenomenon happened due to the breach of sampling balance.

Example 29 (False positives)**.** A rare disease affects one person in 10^5. A test for the disease shows positive with probability 99/100 when applied to an ill person, and with probability 1/100 when applied to a healthy person.

Q: What is the probability that you have the disease given that the test shows positive?

A: Using the Bayes rule, the probability is

$$\begin{aligned} P(\text{ill} \mid +) &= \frac{P(+ \mid \text{ill})P(\text{ill})}{P(+ \mid \text{ill})P(\text{ill}) + P(+ \mid \text{healthy})P(\text{healthy})} \\ &= \frac{0.99\times 10^{-5}}{0.99\times 10^{-5} + 0.01\times(1-10^{-5})} = \frac{99}{99+10^5-1} \\ &= 0.000989 \approx \frac{1}{1011}. \end{aligned}$$

The real chance of ill is rather smaller than intuition. Most positive results are wrong.

Exercise 10. Do the same computation in the above false positive example with probability 10^{-7} of being positive when applied to a healthy person.

Exercises. (DS) 1.2.1, 1.2.4, 1.3.1, 1.3.4, 1.3.5; 1.4.5, 1.4.6, 1.4.9, 1.4.10, 1.4.14, 1.5.7, 1.5.9, 1.5.12, 1.5.13, 1.5.14; 1.6.3, 1.6.8; 1.7.6, 1.7.8, 1.7.11; 1.8.7, 1.8.9, 1.8.14, 1.8.15, 1.8.17, 1.8.22, 1.9.4, 1.9.6, 1.9.10, 1.10.4, 1.10.7, 1.10.12, 1.10.13, 1.12.7, 1.12.12, 1.12.13; 2.1.6, 2.1.11, 2.1.12, 2.1.14, 2.2.4, 2.2.10, 2.2.11, 2.2.16, 2.2.17, 2.2.20, 2.3.1, 2.3.7, 2.3.8, 2.3.16, 2.5.1, 2.5.3, 2.5.6, 2.5.12, 2.5.15, 2.5.17, 2.5.24, 2.5.30; (Rf) P 1.4, 1.6, 1.14, 1.17, 1.30; T 1.8, 1.9, 1.11, 1.12; S 1.9, 1.14, 1.17; (Ri) 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 1.10, 1.12, 1.16, 1.21, 1.23, 1.24, 1.25, 1.26, 1.29, 1.32, 1.43, 1.47, 1.48; (RM) 1.6, 1.7, 1.8, 1.9, 1.11, 1.13, 1.16, 1.21, 1.23, 1.27, 1.34, 1.35, 1.37, 1.42, 1.45, 1.49, 1.54, 1.57, 1.60, 1.64, 1.68, 1.69, 1.70, 1.73, 1.77; (TU) 7.1, 7.4, 7.7, 7.9, 7.10, 7.11, 7.21, 7.23, 8.4, 8.6, 8.7;

References

DS: DeGroot and Schervish

GS: Grimmet and Strikezar

Rf: Ross, first course in probability

Ri: Ross, introduction to probability models

RM: Rice, mathematical statistics and data analysis

TU: Tjims, understanding probability