Statistics (Math 323) winter 2019

Prof: David Wolfson
Yimeng Hu

Contents

1 Lec 01, Jan 08

1.1 Introduction

Why study probability?

- as a discipline in its own right.
- as a part of mathematics/ applied mathematics
- Most importantly, as a tool for statistical inference

The meaning of probability (i.e when we say "the probability" of an event A is 2/3, what do we mean)

$\boxed{\text{Ex}}$ a box has 6 Red and 4 green marbles. Draw a marble at random from the box. What is the probability that the marble is red?

1. If we say P(red) = $\frac{6}{10}$ what do we mean by this statement?

   $\boxed{\textbf{Sol}}$.
   We cannot simply define "the probability" of getting a red as $\lim_{N\to\infty} \frac{\#\ \textit{of red}}{\#\ \textit{of trials}}$ *since we do not know if this limit will exist and be unique for every sequence of trials. So instead in the 1930. The Great Russian mathematician A.N. Komolgorov proposed* ***3 axioms/assumptions*** *that probability should satisfies and then developped a theory of probability from these.*

   Note. *As a consequence of The law of large number, we interpret probability as a limiting relative frequency. However it has nothing to do with relative frequencies.*

2. How did you arrive at this answer?

   $\boxed{\textbf{Sol}}$.
   In order to arrive at the answer $\frac{6}{10}$*, we need to use the Komolgorov axioms and any theorm that follows from that to prove that this is indeed correct.*

1.2 Basic Set Algebra

1. $A = \{\omega : \omega \in A\}$ where A is an event (a set) consist of elementary outcomes w.
2. $A \cap B = \{\omega : \omega \in A \text{and} \omega \in B\}$ **note:** "and" $\Rightarrow$ intersection
3. $A \cup B = \{\omega : \omega \in A \text{or} \omega \in B\}$ **note:** "or" $\Rightarrow$ union.
4. $A \subset B := \omega \in A \Rightarrow \omega \in B$
5. All discussion take place in the context of the **universal set** S
6. $A^c := \{\omega \in S : \omega \notin A\}$ **note:** It is sometimes easier to first find $P(A^c)$
7. $A \cap B = \phi \Rightarrow$ A and B disjoint or mutually exclusive
8. De Morgan's Law:
   a) $(A \cap B)^c = A^c \cup B^c$
   b) $(A \cup B)^c = A^c \cap B^c$

2 Lec 02, Jan 10

2.1 Experiment

$\boxed{\textbf{DEF}}$ **2.1.** *An **experiment** is defined informally as the performance of some actions*

$\boxed{\textbf{DEF}}$ **2.2.** *An **Random Experiment** is one for which the outcome are not known in advance. i.e there is uncertainty in the outcome that will be observed. (Once thet experiment has been conducted though you may not know the outcome, there is nothing random about the outcome)*

$\boxed{\textbf{Ex}}$ **2.1.** *Toss a coin twice and observe the outcome The pre-experiment outcomes are random / uncertain*

$\boxed{\textbf{Ex}}$ **2.2.** *Take 60 subjects who will undergo surgery for a certain disorder before we observe their* $\underbrace{\textit{time to recovery}}_{\textit{outcome}}$, *are random/uncertain*

$\boxed{\textbf{Ex}}$ **2.3.** *Toss a coin until you observe the first head. Let the trial at which this happens be the outcome of interest. This outcome is random before you start tossing.*

2.2 Sample Space

$\boxed{\textbf{DEF}}$ **2.3.** *The set of all possible outcome of an experiment is called the **Sample Space (S)** of the experiment. We denote each outcome as ω, an elemenatry outcome.*

Note. *a sample space mainly depend on how you define your outcomes.*

$\boxed{\textbf{Ex}}$ **2.4.** *Draw a marble at random from 6 Red and 4 Green.*

1. *If order **does not** matter, Let w_1:= event which a red marble is drawn, w_2 := event which a green marble is drawn. then the sample space is as following: $S = \{w_1, w_2\}$*
2. *If order matters, number the marbles WLOG $\{\underbrace{1, ..., 6,}_{Red}\underbrace{, 7, ..., 10}_{Green}\}$ Let w_i := event which marble i is drawn for i = 1,...,10 Then the sample space is: $S = \{w_1, ..., w_{10}\}$*

$\boxed{\textbf{Ex}}$ **2.5.** *Suppose there are n people in a room, ask these people when their birthday are. Let S be the set of outcomes that we could get at the completion of our experiment. Then S could be defined as:*
$S = \{\underbrace{\{Jan1, ..., Jan1\}}_{w_1}, ..., \underbrace{\{Dec31, ..., Dec31\}}_{w_n}\}$

Note. *Here the elementary outcomes are n-dim vectors*

$\boxed{\textbf{Ex}}$ **2.6.** *Toss a coin until you observe the 1^{st} head. Let n be the trial number at which this occurs. Then $S = \{w_1, ..., w_n\}$ = {1,...,n}. S is an example of sample space with a countably many numbers of possible outcomes.*

$\boxed{\textbf{Ex}}$ **2.7.** *Suppose that you measure the height of a dam every July 1^st. The set of possible heights might be S = {[0,20]} where 20 is the height of the dam wall in meters. Here S is an uncountably infinite set. **Note:** In real lift no such sample space exits.*

2.3 Kolmogorov Axioms

$\boxed{\textbf{DEF}}$ **2.4.** *Any subset $E \subset S$ is defined as an* ***event****. The empty set $\phi \subset S$ is also an event.*

$\boxed{\textbf{DEF}}$ **2.5.** *a function P() is a set function on the subset of S if P(A) is a real number for every subset of S. Let S be a sample space. Then P() is called a probability measure if P is a real valued set function on S s.t*

1. $\forall E \subset S, P(E) \geq 0$
2. *P(S) = 1*
3. *Let $E_1, E_2, ...$ be any countable connection of events such that they are mutually exclusive, i.e $E_i \cap E_j = \phi \qquad \forall i \neq j$. Then $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$*

Note. *From these 3 axioms, we developpe the entire theory of probability including the Law of Large numbers which allows us to interpret probability as a limiting relative frequency. We shall state and prove 5 theorem that will be useful for solving word problems.*

3 Lec 03, Jan 15

3.1 The 5 theorems

Thm 3.1. 1. *For any event A, $P(A^c)$ = 1-P(A)*

<u>*PROOF:*</u>.

$$\begin{aligned} A \cup A^c = S, \Rightarrow P(A \cup A^c) = P(S) = 1 \quad &\text{(Ax 2)} \\ A \cap A^c = \phi \Rightarrow P(A \cup A^c) = P(A) + P(A^c) \quad &\text{(Ax 3)} \\ \Rightarrow P(A^c) = 1 - P(A) \end{aligned}$$

□

2. *$P(\phi) = 0$*
3. *$P(A \cap B^c) = P(A) - P(A \cap B)$*

<u>*PROOF:*</u>. trick: Try to write unions as disjoint unions and apply Ax 3

$$\begin{aligned} A = (A \cap B) \cup (A \cap B^c) &\Rightarrow P(A) = P(A \cap B) + P(A \cap B^c) \quad \text{(Ax 3)} \\ &\Rightarrow P(A \cap B^c) = P(A) - P(A \cap B) \end{aligned}$$

□

4. *$A \subset B \Rightarrow P(A) \leq P(B)$*

<u>*PROOF:*</u>.

$$\begin{aligned} B = A \cup (B \cap A^c), A \cap (B \cap A^c) = \phi \\ \Rightarrow P(B) &= P(A) + P(B \cap A^c) \\ &\geq P(A) \quad \text{(Ax 1)} \end{aligned}$$

□

5. *For any two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$*

Note. *if $A \cap B = \phi$, then $P(A \cup B) = P(A) + P(B)$*

<u>*PROOF:*</u>.

$$A \cup B = \underbrace{(A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)}_{\text{mutually exclusive}}$$

$$\begin{aligned} \Rightarrow P(A \cup B) &= P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) \quad &\text{(Ax 3)} \\ &= P(A) - P(A \cap B) + P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \quad &\text{(Thm 3)} \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

□

Cor 3.1.1. *For any event A, $0 \leq P(A) \leq 1$*

Note. *Do not use tree diagram to present your answer. Start by defining the simplest possible event then construct more complicated event by using set operations.*

"Either ... or...", "At least" $\Rightarrow \bigcup$, *"and"* $\Rightarrow \bigcap$, *"Not"* $\Rightarrow$ *complement.*

Then apply axiom or theorem.

Ex **3.1.** *Suppose that it's known that 20% of people smoke and that 1% of old people will developpe lung cancer. Suppose that the probability of someone will either smoke or developpe lung cancer is 0.205. Let A:=the event of someone smokes.*
B:= the event of someone has cancer
Then $P(A) = 20\%$, $P(B) = 1\%$ *and* $P(A \cup B) = 0.205$

1. *Find the proportion of people who smoke and developpe lung cancer.*

   Sol. *WTS* $P(A \cap B)$

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) && \text{(thm 5)} \\ \Rightarrow P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0.2 + 0.01 - 0.205 \\ &= 0.005 \end{aligned}$$

2. *What is the probability that someone does not smoke but have lung cancer.*

   Sol. *WTS* $P(A^c \cap B)$

$$\begin{aligned} P(A^c \cap B) &= P(B) - P(A \cap B) && \text{(thm 3)} \\ &= 0.01 - 0.005 \\ &= 0.005 \end{aligned}$$

3. *What is the probability that someone smokes but does not have lung cancer.*

   Sol. *WTS* $P(A \cap B^c)$

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) && \text{(thm 3)} \\ &= 0.2 - 0.005 \\ &= 0.195 \end{aligned}$$

4. *What is the probability that someone neither smoke nor have lung cancer.*

   Sol. *WTS* $P(A^c \cap B^c)$

$$\begin{aligned} P(A^c \cap B^c) &= P((A \cup B)^c) && \text{(De Morgan's Law)} \\ &= 1 - P(A \cup B) && \text{(thm 1)} \\ &= 1 - 0.205 \\ &= 0.195 \end{aligned}$$

3.2 Tools for calculating probability

Thm **3.2.** *Let S be a finit sample space with N equally likely outcomes. Let E be any event in S. Then*

$$P(E) = \frac{\mid E \mid}{N} = \frac{\text{\# of outcoms in } E}{Tot.possibleoutcomes}$$

Note. *The calculation of a probability can be then reduced to a counting problem*

PROOF:. write the event E as the union of the elementary outcomes i.e

$$E = \bigcup_{w_i \in E} w_i \Rightarrow P(E) = P(\bigcup_{w_i \in E} w_i) = \sum_{w_i \in E} P(W_i)$$

$$P(S) = \sum_{i=1}^{N} P(w_i) = 1 \Rightarrow P(w_i) = \frac{1}{N} \qquad \forall i = 1, ..., N \tag{Ax 2}$$

Hence:
$$\begin{aligned} P(E) &= \sum_{i=w_i \in E} \frac{1}{N} \\ &= \frac{1}{N} \sum_{i=w_i \in E} 1 \\ &= \frac{\mid E \mid}{N} \end{aligned}$$

□

4 Lec 04, Jan 17

we want a sample space with equally likely outcomes.

Recall. *$S = \{1,2, \dots , 10\}$, $R = \{1,2,3,4,5,6\}$, $G = \{7,8,9,10\}$*
All of these outcomes are reasonably equally likely. There are N=10 such outcomes.
Therefore by the above thm(last class), $P(R) = \frac{\text{\# of ways to get a red marble}}{\text{tot. no. of possible outcomes}} = \frac{6}{10}$.

Note. *Now although the above thm(last class) is easy to understand, the counting can sometimes be very difficult. it is useful to have some counting tools*

4.1 Counting Rule

1. If you have a set of n distinct object, then the number of ways to order the objects in n!
2. If you have a set of n distinct object, then the number of ways to draw r object from the set, and the order is unimportant, sampling **without** replacement is denoted as "n choose r".
$$\binom{n}{r} = \frac{n!}{(n-r)!r!}, \qquad \text{note: } 0! = 1 \text{ by def}$$
3. If we have a set of n distinct objects. The number of ways to draw r objects from these n and the order does matter, sampling **without** replacement is denoted by "n permutation r"
$$P(n,r) = \frac{n!}{(n-r)!}$$
4. **Multiplication Rule** (Sausage Rule)
Suppose that you have k set of $n_1, n_2, ..., n_k$ distinct objects respectively. The number of ways to form a set by selecting one object from each set is given by
$$n_1 * n_2 * ... * n_k$$

4.2 The Birthday Problem

Suppose there are n people in a room. What is the probability that at least two have the same birthday?

PROOF:.
Suppose that there are 356 possbile birthday
Let E := event that at least two people have the same birthday.
It is easier to compute $P(E^c)$ = P(no two have the same birthday), then
$$P(E) = 1 - P(E^c) \qquad \text{(by Thm 1)}$$
The sample space here is S = {(Jan 1, ..., Jan 1), ... , (Dec 31, ..., Dec 31)}
First we assume that all of these outcomes are equally likely. There are finitely many of them. Therefore
$$\begin{aligned} P(E^c) &= \frac{\text{\# of ways that } E^c \text{can occur}}{Tot.no.of outcomes ins} \\ &= \frac{\text{P(365,n)}}{365^n} \\ &= \frac{365 * 364 * ... * (365 - n + 1)}{365^n} \end{aligned}$$
Hence
$$P(E) = 1 - \frac{365 * 364 * ... * (365 - n + 1)}{365^n}$$
□

4.3 The Fish in the Lake Problem

Suppose a lake has N fish in it, of which **a** are tagged and **N-a** are untagged. If you draw a sample n fish from the lake, sampling **without** replacement, What is the probability of getting **x** tagged fish in my sample?

PROOF:.
We want a sample space with equally likely outcomes.
Start by numbering the fish from 1 to N.
We will suppose that the fish with numbers 1,...,**a** correspond to those with tags and the remaining **N-a** numbers to the untagged fish.
An outcome for our experiment is defined to a set of n integers selected from the integers 1,...,N
The order is considered unimportant and assume that all sets of n numbers are equally likely.
Hence we can use our thm to solve the problem.
Let E:=event that there are x tagged in sample

$$P(E) = \frac{\text{number of ways to get x tagged}}{\text{total number of poss. outcomes}}$$

We have

$$\begin{aligned}\text{Tot. number of possible outcomes} &= \text{number of ways to draw n integers from a set of N distinct integers} \\ &= \binom{N}{n} \qquad \text{(counting rule 2.)}\end{aligned}$$

Now use **Multiplication Rule**
each sub-sausage contains $x \leq a$ integers selected from integer 1,...,a
each sub-sausage contains n-x integers selected from integers (a+1),...,N. So we must count the number of objects in each of these two sausages, n_1, n_2. say for the number of ways to get x tagged fish is $= n_1 * x * n_2$
we have

$$n_1 = \binom{a}{x}, \qquad n_2 = \binom{N-a}{n-x}$$

finally

$$P(\text{x tagged fish out of n}) = \frac{\binom{a}{x} * \binom{N-a}{n-x}}{\binom{N}{n}}$$

□

4.4 Capture & Recapture Problem

Have N fish in the lake ***Capture Phase***

1. Remove and tag ***a*** fish.
2. Return the fish to the lake

Recapture Phase

1. capture ***n*** fish
2. Count how many tagged fish in the recaptured sample

$$P(X = x) = \frac{\binom{a}{x} * \binom{N-a}{n-x}}{\binom{N}{n}}$$

N is unknown, But

$$\frac{a}{N} \approx \frac{x}{n}$$
$$N = \frac{a * n}{x}$$

However, in real world, the captured face tends to be harder to be recaptured

5 Lec 05, Jan 22

5.1 Conditional Probability

Idea: Sometimes, knowning that an event A has occured influences the probability that the event B will occure.

Ex 5.1. *In our marble problem, The probability of getting a red marble on the second of two draws (**without** replacement) knowning that we got a red on the first, is different from simply the probability of getting a red on the second draw. Argument*

$$\begin{aligned} P(R_2) &= P[(R_2 \cap G_1) \cup (R_2 \cap R_1)] \\ &= P[R_2 \cap G_1] + P[R_2 \cap R_1] \end{aligned}$$

which is easy to see, $P(R_2)$ is different from $P(R_2$ knowning $R_1)$

We therefore feel justified in formally defining the notion of "Conditional Probability"

DEF 5.1. *Let A and B be two events such that P(A) $\neq$ 0, then we define the probability of B given A as follows:*

$$P[B \text{ given } A] := P[B \mid A] = \frac{P[A \cap B]}{P(A)}$$

Note: *the RHS is the ratio of two probability and we have defined probability (The 3 Axioms)*

Note.

1. *we need to check whether conditional probability satisfies the 3 Axioms.*

   (a) $P(B \mid A) \geq 0$

   <u>*PROOF:*</u>.

   $$P[B \mid A] = \frac{\underbrace{P[A \cap B]}_{\geq} 0}{\underbrace{P(A))}_{\geq}}$$

   Hence true □

   (b) $P[S \mid A] = 1$

   <u>*PROOF:*</u>.

   $$P[S \mid A] = \frac{P[A \cap S]}{P(A)} = \frac{P(A)}{P(A)} = 1$$

   Hence true □

   (c) $P\left[\bigcup B_i \mid A\right] = \sum_{i=1}^{\infty} P[B_i \mid A] \qquad where B_i \cap B_j = \phi \qquad \forall i \neq j$

   <u>*PROOF:*</u>. (exercise) □

   It then follows that the 5 theorem also go through for conditional Probability

2. $P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad$ *if* $P(B) \neq 0$

The definition of conditional probability leads to a fundamental theorem that allows us to sometimes find the probability of an intersection

5.2 Multiplication Rule for Conditional Probability

Follows immediately from the definition of conditional probability

$$\begin{aligned} P(B \cap A) &= P(B \mid A) * P(A) \\ &= P(A \mid B) * P(B) \end{aligned}$$

The hope is that when you are required to find $P[A \cap B]$, you know either P(A) or P(B) and one of the conditional probability.
In word problems,

$$\text{"of those that"} \Rightarrow \text{Conditional Probability}$$

Do not confuse "and" with "given that"

$\boxed{\textbf{Ex}}$ **5.2.** *We have two inspectors for items coming off an assembly . The proportion of items that are declared non-defective by the first inspector is 0.90 of those items that are declared non-defective by inspector 1, 0.95 are declared non-defective by inspector 2. What is the probability that an item is declared non-defective by both inspectors.*

$\boxed{\textbf{Sol}}$.

Let ND_i (i=1,2) := event non-defective for each of the inspectors resp.
WTS: $P\big[ND_1 \cap ND_2\big]$
Given: $P[ND_1] = 0.90$ *and* $P\big[ND_2 \mid ND_1\big] = 0.95$

$$\begin{aligned} P[ND_2 \cap ND_2] &= P[ND_2 \mid ND_1] * P[ND_1] \\ &= 0.90 * 0.95 \end{aligned}$$

□

***Extension:** LEt $A_1, ..., A_n$ be any sequence of events. Then

$$\begin{aligned} P[A_1 \cap ... \cap A_n] &= P[A_n \mid A_1 \cap ... \cap A_{n-1}] * P[A_1 \cap ... \cap A_{n-1}] \\ &= P[A_n \mid A_1 \cap ... \cap A_{n-1}] * P[A_{n-1} \mid A_1 \cap ... \cap A_{n-2}] * P[A_1 \cap ... \cap A_{n-2}] \\ &= ... \\ &= \Pi_{i=1}^{n} P[A_i \mid A_1 \cap ... \cap A_{i-1}] * P(A_0) \qquad \text{(i=1,...,n)} \end{aligned}$$

$\boxed{\textbf{Ex}}$ **5.3.**

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3|A_1 \cap A_2)P(A_1 \cap A_2) \\ &= P(A_3|A_1 \cap A_2)P(A_2|A_1)P(A_1) \end{aligned}$$

The process of repeatedly conditioning stating with the last event is called the process of ***conditioning backwards***.
When you are required to find the probability of the intersection of several evetns, think of conditioning backwards

6 Lec 06, Jan 24

6.1 Conditioning Backwards

DEF **6.1.** *The process of repeatedly conditioning stating with the last event is called the process of **conditioning backwards**.*
When you are required to find the probability of the intersection of several evetns, think of conditioning backwards

Ex **6.1.** ***The Marble Problem***

1. *Suppose that you draw 2 marbles **without** replacement. What is the probability that the second marble drawn is green?*

   Sol. *Conditioning Backwards.*

   $$G_2 = (R_1 \cap G_2) \cup (G_1 \cap G_2)$$

   Implies

   $$\begin{aligned} P(G_2) &= P((R_1 \cap G_2)) + P((G_1 \cap G_2)) && \text{(Ax 3)} \\ &= P(G_2|R_1)P(R_1) + P(G_2|G_1)P(G_1) \\ &= \frac{4}{9} * \frac{6}{10} + \frac{3}{9} * \frac{4}{10} \end{aligned}$$

2. *Suppose that you draw 5 marbles. What is the probability that you will get the sequence R_1, R_2, G_3, G_4, R_5.*

   Sol. *Conditioning Backwards*

   $$\begin{aligned} P(R_1 \cap R_2 \cap G_3 \cap G_4 \cap R_5) = P(R_5|R_1 \cap R_2 \cap G_3 \cap G_4) * P(G_4|R_1 \cap R_2 \cap G_3) \\ * (P(G_3|R_1 \cap R_2) * (P(R_2|R_1) * P(R_1) \end{aligned}$$

   Hence

   $$P(R_1 \cap R_2 \cap G_3 \cap G_4 \cap R_5) = \frac{4}{6} * \frac{3}{7} * \frac{4}{8} * \frac{5}{9} * \frac{6}{10}$$

The following theorem on conditional probability are fundamental

6.2 The Law of Total Probability

Thm **6.1.**
Let A be any event, let $B_1, B_2, ...$ be m events that satisfy the following

1. $B_i \cap B_j = \phi \qquad \forall i \neq j$
2. $\bigcup_{i=1}^{m} B_i = S$ *we call* $\{B_1, B_2, ...\}$ *a partition of* S

Then $P(A) = \sum_{i=1}^{m} P(A|B_i)P(B_i)$

<u>*PROOF:*</u>. (Of theorem)
known $A = \bigcup_{i=1}^{m} \underbrace{(A \cap B_i)}_{\text{all disjoint}}$ Hence

$$\begin{aligned} P(A) &= \sum_{i=1}^{m} P(A \cap B_i) && \text{(Ax 3)} \\ &= \sum_{i=1}^{m} P(A|B_i)P(B_i) \end{aligned}$$

□

Note. *Maybe A is complicated and it is difficult to find its probability directly or the given information does not provide P(A) directly.*
The hope is that we can find $P(A|B_i)$ easily or that they come with the provided information and that we know $P(B_i)$.
In word problem, the clue to use the Law of Total Probability is that you are given a bunch of conditional probability and the probability $P(B_i)$ and you are asked to find P(A).

6.3 Baye's Theorem

Thm **6.2.** *Let A and $B_1, B_2, ...$ be defined exactly as in the Law of Probability. Then we can write*

$$P(B_k|A) = \frac{P(A|B_k) * P(B_k)}{\sum_{i=1}^{m} P(A|B_i)P(B_i)} \qquad (k = 1, 2, ..., m)$$

<u>*PROOF:*</u>.

$$\begin{aligned} P(B_k|A) &= \frac{P(B_k \cap A)}{P(A)} \\ &= \frac{\overbrace{P(A|B_k)P(B_k)}^{\text{mult. rule}}}{\underbrace{\sum_{i=1}^{m} P(A|B_1)P(B_i)}_{\text{Law of tot.Prob.}}} \end{aligned}$$

□

Note. *Mathematically, Baye's Theorem allows you to reverse one or more given conditional probability.*
In word problem, the clue to use the Baye's thm is that you are required to reverse one or more conditional probability statement.

$\boxed{\textbf{Ex}}$ **6.2.** *Suppose that there is a diagnostic test for breast cancer and that in a certain population $\frac{5}{1000}$ women have breast cancer. Known that the test has the following properties:*

1. *if a woman has breast cancer, the test will be positive 95% of the time.*
2. *if a woman does not have breast cancer, the test will be negative 95% of the time.*

Question is

i) *What proportion of women will test positive?*

$\boxed{\textbf{Sol}}$.

$$\begin{aligned} \text{Let } Pos &:= \text{event that a test is positive} \\ \text{Let } Neg &:= \text{event that a test is negative} \\ \text{Let } Bc &:= \text{event that a woman has breast cancer} \\ \text{Let } Bc^c &:= \text{event that a woman does not have breast cancer} \end{aligned}$$

Known

$$\begin{aligned} P(Pos\ |Bc) &= 95\% \\ P(Pos\ |Bc^c) &= 1 - 95\% = 0.05 \\ P(Neg\ |Bc^c) &= 95\% \\ P(BC) &= 0.005 \\ P(BC^c) &= 1 - 0.005 \end{aligned}$$

We have that BC and BC^c *are disjoint and* $BC \cup BC^c = S$.
Therefore by the Law of Tot. Prob.

$$\begin{aligned} P(\ Pos\ ) &= P(Pos\ |Bc) * P(BC) + P(P(Pos\ |Bc^c)) * P(Bc^c)) \\ &= 0.95 * 0.005 + 0.05 * (1 - 0.005) \\ &= 0.054 \end{aligned}$$

ii) *If a woman tests postive, what is the probability that she has breast cancer?*

$\boxed{\textbf{Sol}}$. *WTS P(Bc | Pos).*
note: *we are required to reverse the P(Pos | Bc) (Baye's)*

$$\begin{aligned} P(Bc\ |Pos\ ) &= \frac{P(Pos\ |Bc) * P(Bc)}{P(Pos\ |Bc) * P(Bc) + P(Pos\ |Bc^c) * P(Bc^c)} \\ &= \frac{0.95 * 0.005}{\underbrace{0.054}_{\text{from part i)}}} \\ &= 0.087 \end{aligned}$$

Some comments on Baye's Theorem and diagnostic tests:
In the world of diagnostic tests P(Pos | Disease) is called the sensitivity of the test P(neg |Disease^c) is called the specificity. P(Disease) is called the disease prevelance and P(Disease | Pos) is called the postive predictive value of the test. Note that pos predictive value depends on the sensitivity, specitivity and prevalue of the diseasea.

7 Lec 07, Jan 29

Some comments on Baye's Theorem and diagnostic tests:
In the world of diagnostic tests **P(Pos | Disease)**:= the sensitivity of the test
P(neg |Disease^c):= the specificity.
P(Disease):= the disease prevelance and
P(Disease | Pos):= the postive predictive value of the test.
Note that pos predictive value depends on the sensitivity, specitivity and pre. value of the diseasea.

7.1 Statistical Independence*

Idea: sometimes the occurance of an event A does **NOT** influence the probability that an event B will occur.

$\boxed{\textbf{Ex}}$ **7.1.** *In the marble problem, supose that you draw two marbles* ***with*** *replacement. What is the probability that the second marble is drawn is red given the first is green. (i.e what is $P[R_2|G_1]$).*
Clearly, we obtained a green on the first draw has no impact on the probability of a red on the second draw, since the box was returned to its original composition. This idea leads to a definition of the independence of two event A and B

$\boxed{\textbf{DEF}}$ **7.1.**

1. *We say that the events A and B are independent $\iff$ P[B|A] = P(B)*
   Note: *This definition is while intuitive, it is not easily extended to the notion of independence of more than two events.*

2. *(non-intuitive but extendable) The events A and B are said to be independent,*
   $\iff$ $P(A \cap B) = P(A) * P(B)$.

$\boxed{\textbf{Thm}}$ **7.1.** *A and B are independent according to DEF 1 $\iff$ they are independent according to DEF 2. (i.e The above two definition are equivalent.)*

<u>*PROOF:*</u>.

(⇒) Assume A and B are independent according to DEF 1, WTS they are also independent according to DEF2.
Then P(B | A) = P(B).
Hence by the Multiplication rule

$$\begin{aligned} P(A \cap B) &= P(B|A) * P(A) \\ &= P(B) * P(A) \end{aligned}$$

(⇐) Trivial

□

$\boxed{\textbf{DEF}}$ **7.2.** *The events $A_1, A_2, ..., A_n$ are sait to be mutually independent if $P[A_{i_1} \cap ... \cap A_{i_k}] = P[A_{i_1}] * ... * P[A_{i_k}]$ for all subset of $A_{i_1}...A_{i_k}$ selected from A_1 to A_n.*

$\boxed{\textbf{Ex}}$ **7.2.**

1. $P(A \cap B \cap C) = P(A)P(B)P(C)$

2. $P(A \cap B) = P(A)P(B)$ *etc...*

We also define independence for an infinite sequence of events $A_1, A_2, ...A_3$

$\boxed{\textbf{DEF}}$ **7.3.** *The event $A_1, A_2, ...$ are independent if and only if every finite set of A_i is independent according to the previous definition of independence.*

Note.

1. *Events $A_1, A_2, ..., A_n$ are said to be pairwise independent if $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $i \neq j$. It can be shown that pairwise independence does **NOT** imply mutual independece.*
2. *we write $A \perp\!\!\!\perp B$ to denote A is independent of B*
3. *If $A \perp\!\!\!\perp B$ then $B \perp\!\!\!\perp A$ and vice-versa.*
   Further, we have if A, B and C are mutually independent then
   (a) $A^c \perp\!\!\!\perp B$
   (b) $(A \cup B)^c \perp\!\!\!\perp C$
   (c) $(A^c \cap B^c) \perp\!\!\!\perp C$
   (d) $(A \cup C)^c \perp\!\!\!\perp B$
   (e) $A^c \perp\!\!\!\perp B^c$ *etc...*

   PROOF:. WTS $\mathrm{P}(A^c \cap B^c) = \mathrm{P}(A^c)\mathrm{P}(B^c)$

$$\begin{aligned}
P(A^c \cap B^c) &= P((A \cup B)^c) && \text{(De Morgan)} \\
&= 1 - P(A \cup B) && \text{(Thm 1)} \\
&= 1 - \Big[P(A) + P(B) - P(A \cap B)\Big] && \text{(Thm 5)} \\
&= 1 - P(A) - P(B) + P(A \cap B)
\end{aligned}$$

   Now notice

$$\begin{aligned}
P(A^c)P(B^c) &= (1 - P(A))(1 - P(B)) \\
&= 1 - P(A) - P(B) + P(A)P(B)
\end{aligned}$$

   Hence

$$P(A^c \cap B^c) = P(A^c)P(B^c)$$

   □

4. *In fact the following is true*
   Suppose that you have two sets $A_1, A_2, ...A_n$ and $B_1, B_2, ..., B_n$. We say that the set of A is independent of set B if the probability of the intersection of every set of A with the intersection of every set of B is the product of the intersection of the set of A and the set of B.
   E.g $P\Big[(A_3 \cap A_5 \cap A_6) \cap (B_1 \cap B_2)\Big] = P\Big[A_3 \cap A_5 \cap A_6\Big] * P\Big[B_1 \cap B_2\Big]$

8 Lec 08, Jan 31

8.1 The Role of Independence

1. The most important Role of Independence in probability is the following:
   If you can assume independece base on your knowledge of the substantive area and/or the way the experiment was carried out. Then subsequent probability calculations often become a lot easier than if you cannot make this assumptions.
   This is so since the probability of intersection. becomes product of probabilities rather than product of conditional probabilities requiring knowledge of these conditional probabilities.

2. Second, we may want to decide whether evetns are independent (i.e This may be the goal.)
   For instance, one may wish to know, whether recovery time from abdominal surgery is independent of the temperature of the operating room.

3. The relation between disjointness and independence:
   It turns out that these notions are completely different. Disjointness is entirely a set property whereas independece depends on how probabilities are assigned to these events. The following theorem says it all:

   $\boxed{\textbf{Thm}}$ **8.1.** *Suppose that A and B are disjoint, then A and B are independent only if either P(A) = 0, or P(B) = 0.*

   PROOF:.

   $$\begin{aligned}\text{A, B disjoint} &\Rightarrow A \cap B = \phi \\ &\Rightarrow P(A \cap B) = P(\phi) = 0\end{aligned}$$

   Now if A $\perp\!\!\!\perp$ B, then we must have:

   $$\begin{aligned}P(A \cap B) = P(A)P(B) &\Rightarrow P(A)P(B) = 0 \\ &\Rightarrow P(A) = 0, \text{ or } P(B) = 0\end{aligned}$$

   $\square$

   Note. *sometimes you will be required to find $P(A \cup B)$, you have from theorem 5, that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,*
   To deal with $P(A \cap B)$:

   (a) $A \cap B = \phi \Rightarrow P(A \cap B) = 0$
   (b) $A \perp\!\!\!\perp B \Rightarrow P(A \cap B) = P(A)P(B)$
   (c) A and B are dependent, then $P(A \cap B) = P(B \mid A)P(A)$ (note, this is always true)

Ex **8.1.** *(on how independence can be used)*

Note. *Preliminary note*

1. *when sampling* ***without*** *replacement, the outcomes in the sequence of draws are dependent.*
   Thus if you have a box of 10 items if which 4 are defective and you remove 2 without replacement, whether or not you observe a defective on the second draw will depend on what was removed on the first draw. However, if "the box" from which we sample is very large, relative to the size of the sample, we may regard the outcomes of our draws, as being roughly independent.

Suppose that in a very large city, 20% of people have a certain genetic mutation. If 10 people are examined
Clearly we are sampling without replacement (by design).
Let the outcome on trial i, be M_i:= there is a mutation for subjects i.
M_i^c:= there is no mutation for subjects i.
Let X:= the number of mutations in these ten trials. (randome variable.)

1. *What is the probability that exactly 2 will have the mutation?*

   Sol. *WTS P(X = 2). Proceed as following*

   (a) *Find the probability of a specific configuration of mutations and non-mutations in 10 trials that result in 2 mutations out of 10.*
   Consider the configuration

$$
\begin{aligned}
(M_1, M_2, M_3^c, ..., M_{10}^c) &= M_1 \cap M_2 \cap M_3^c \cap ... \cap M_{10}^c \\
P(M_1, M_2, M_3^c, ..., M_{10}^c) &= P(M_1)P(M_2)P(M_3^c)...P(M_{10}^c) \\
&\quad \text{(Rough independenc (large 'city' - small sample))} \\
&= 0.2 * 0.2 * 0.8 * ... * 0.8
\end{aligned}
$$

   Now notice, all configurations which result in exactly two mutations will have probability $(0.2)^2(0.8)^8$

   (b) *Sum up the probability of all such configurations*

$$
\begin{aligned}
P[X = 2] &= P\Big[\bigcup_{k=1}^{\textit{all config.}} \textit{configuration k with 2 mutations} \Big] \\
&= \Sigma_{k=1}^{\textit{all config.}} P\Big[\textit{configuration k with 2 mutations} \Big] && \text{(Ax 3. (config. disjoint))} \\
&= (0.2)^2(0.8)^8 \Sigma_{k=1}^{\textit{all config.}} 1 \\
&= (0.2)^2(0.8)^8 \binom{10}{2}
\end{aligned}
$$

2. *What is the probability that at least 2 will have the mutation?*
   (next time)