ST5201: Basic Statistical Theory
Chap1: Probability

Choi, Yunjin
stachoiy@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

August 22nd, 2017

Outline

- Introduction
- Sample Spaces
- Probability Measures
- Computing Probabilities: Counting Methods
- Conditional Probability
- Independence

Introduction

Learning Outcomes

- Questions to Address:
  What probability is $\star$ Understanding algebra of events $\star$ Various properties of probability $\star$ How to count $\star$ How to compute probability $\star$ Difference between combination & permutation $\star$ What a conditional probability is $\star$ What independence is

Concept & Terminology

- experiment ⋆ sample space ⋆ set theory ⋆ event
- complement/union/intersection ⋆ null/mutually exclusive/disjoint/exhaustive events ⋆ Venn diagram
- commutative/associative/distributive/DeMorgans laws ⋆ probability measure ⋆ addition/multiplication law
- equally likely outcomes ⋆ multiplication principle ⋆ permutation & combination ⋆ conditional probability
- law of total probability ⋆ tree diagram ⋆ Bayes rule
- independence ⋆ pairwise/mutual independence

Mandatory Reading

Textbook: Section 1.1 – Section 1.6

What is Probability?

History

Gambling shows that there has been an interest in quantifying the ideas of probability for millennia, but exact mathematical descriptions arose much later. Jakob Bernoulli's [1](posthumous, 1713) treated the subject as a branch of mathematics.

Probability, in practice

- is the measure of the likeliness that an event will occur.
- The higher the probability of an event, the more certain we are that the event will occur.
- Quantification of the uncertainty appeared in many fields, e.g.
  - genetics & bioinformatics (mutations)
  - operations research (demands on the inventories of goods)
  - finance (volatility of a stock)

[1] Ars Conjectandi

Definition

An *experiment* is any action or process whose outcome is subject to uncertainty/randomness.

- An experiment has multiple outcomes.
- When an experiment is conducted, ONLY one of all possible outcomes would occur. It is uncertain which outcome would occur.
- Assume that the set/collection of all possible outcomes is known

Definition

The *sample space* of an experiment, denoted by Ω, is the set/collection of all possible outcomes of that experiment. Each element of Ω, denoted by ω, is an outcome.

- The simplest experiment is one with 2 possible outcomes, e.g.,[2]

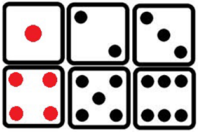
  When we flip/toss a coin & see which side faces up, $\Omega = \{H, T\}$, where H & T denote head facing up & tail facing up, respectively

  In an experiment of examining a light bulb to see if it is defective, $\Omega = \{$, $\}$

[2] https://www.google.com.sg/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=flip+a+coin

Examples-cont'd

When we roll a die & see the upturned #, the sample space is $\Omega = \{1, 2, \ldots, 6\}$



If we examine 3 light bulbs in sequence & note the result of each examination, then an outcome for this experiment is any sequence of 's & 's of length 3, so

$\Omega = \{$  $\}$

The amount of time between successive customers arriving at a check–out counter is of interest. For such an experiment,

$$\Omega = \{t | t \geq 0\}$$

- Interested in not only the individual outcomes of Ω but also various collections of outcomes from Ω

Definition

An *event* A is any collection of outcomes contained in the sample space (i.e., any *subset* of Ω written as $A \subset \Omega$). An event is *simple* if it consists of exactly 1 outcome & *compound* if it consists of > 1 outcome

- When an experiment is performed & an outcome $\omega \in \Omega$ is observed/realized:
  - An event A is said to occur if the observed outcome ω is included in A (i.e., $\omega \in A$)
  - Exactly 1 simple event ($\{\omega\}$) occurs, but many compound events occur simultaneously

- **Toss 2 coins**: $\Omega = \{HH, HT, TH, TT\}$
  Event A: 2 heads are observed
  $A = \{HH\}$, is a simple event as A contains only 1 outcome
  Event B: exactly 1 head is observed
  $B = \{HT, TH\}$, is a compound event
- **Examine 3 light bulbs**: Different compound events inlude, e.g.

A = {, , }

= the event that exactly 1 of the 3 light bulbs is not defective

B = {, , , }

= the event that at most 1 light bulb is defective

C = {, }

= the event that all 3 light bulbs are in the same condition

Suppose that the experiment is performed, & the observed outcome is . Then, the simple event {} has occurred & so have the compound events B & C (but not A)

Algebra of Events

- As an event is just a set, so set operations from elementary set theory carry over directly into prob theory, which allows us to *create new & more "complex" events from given events*

Definition

Given any event $A, B \subset \Omega$,

- The *complement* of A, denoted by A^c, $\bar{A}$, or A', is the set of all outcomes in Ω that are not contained in A
- The *union* of sets A and B, denoted by $A \cup B$ (read "A or B"), is the event consisting of all outcomes that are *either in A or B or in both events*
- The *intersection* of sets A and B, denoted by $A \cap B$ or AB & read "A and B", is the event consisting of all outcomes that are *both in A and B*

Commutative laws: $A \cup B = B \cup A$, $A \cap B = B \cap A$

Definition

The *null event*, denoted by $\emptyset$, is the event *consisting of no outcomes*

Definition

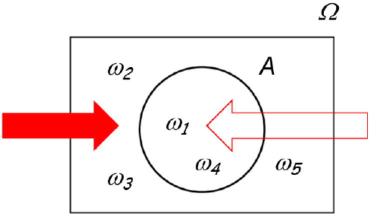
We say A and B are *disjoint* or *mutually exclusive* events when $A \cap B = \emptyset$. It follows that A and A^c must be disjoint for any event $A \subset \Omega$.

Definition

We say A and B are *exhaustive* events when $A \cup B = \Omega$. It follows that A and A^c must be exhaustive for any event $A \subset \Omega$.

Venn diagrams: a useful tool for visualizing set operations

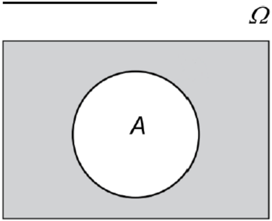
All outcomes $\omega \in \Omega$ are inside the rectangle Ω



All outcomes $\omega \in A$ are inside the circle representing A

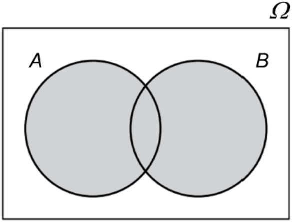
Note: Events can be represented by objects of *any shape*

Complement: A^c is the shaded area

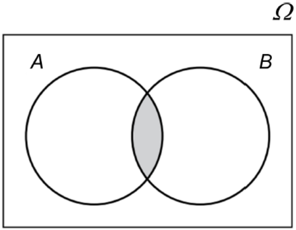


Venn Diagrams-cont'd

Union: $A \cup B$ is the shaded area

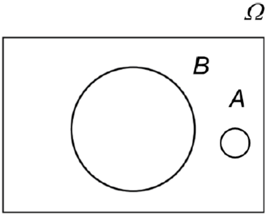


Intersection: $A \cap B$ is the shaded area

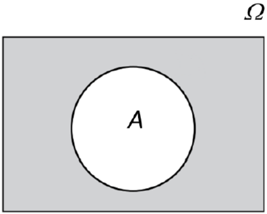


Venn Diagrams-cont'd

A & B are *mutually exclusive or disjoint events*:

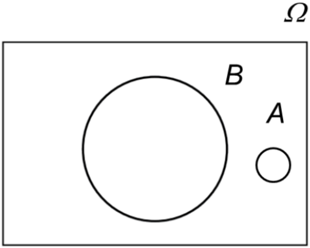


A & A^c (represented by the shaded area) are *mutually exclusive* & *exhaustive events*:



Examples-cont'd

Toss 2 coins: $\Omega = \{HH, HT, TH, TT\}$; $A = \{HH\}$ is the right circle; $B = \{HT, TH\}$ is the left circle



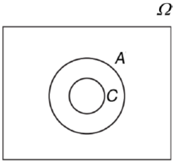
How about $C = \{TT\}$?

Assume that there are 7 possible outcomes in an experiment such that (s.t.)

$$\Omega = \{0, 1, 2, 3, 4, 5, 6\}$$

Let $A = \{0, 1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$ & $C = \{1, 2\}$. Then,

$$\begin{aligned}
A^c &= \{5, 6\} \\
A \cup B &= \{0, 1, 2, 3, 4, 5, 6\} = \Omega \text{ (}i.e.,\ A \text{ \& } B \text{ are } \textit{exhaustive}\text{)} \\
A \cup C &= \{0, 1, 2, 3, 4\} = A \text{ (indeed, } C \text{ is a } \textit{subset} \text{ of } A\text{)} \\
A \cap B &= \{3, 4\} \\
A \cap C &= \{1, 2\} \\
(A \cap C)^c &= \{0, 3, 4, 5, 6\} \\
B \cap C &= \varnothing \text{ (}i.e.,\ B \text{ \& } C \text{ are } \textit{disjoint}\text{)}
\end{aligned}$$



Algebra for Multiple Events

The operations of union & intersection can be extended to ≥ 3 events, and the idea of disjointness & exhaustiveness can also be generalized

For any 3 events, $A, B, C \subset \Omega$

- $A, B \& C$ are said to be *mutually exclusive or pairwise disjoint* if no 2 events have any outcomes in common
- $A, B \& C$ are said to be *exhaustive* if the event $A \cup B \cup C$ consists of all outcomes in Ω

Laws for Multiple Events

Some Useful Laws

Given any event $A, B, C, E_1, \cdots, E_n \subset \Omega$,

- *Associative Laws*:

$$(A \cup B) \cup C = A \cup (B \cup C) \qquad (A \cap B) \cap C = A \cap (B \cap C)$$

- *Distributive Laws*:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \qquad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

- *DeMorgan's Laws*:

$$\left(\bigcup_{i=1}^{n} E_i \right)^c = \bigcap_{i=1}^{n} E_i^c \qquad \left(\bigcap_{i=1}^{n} E_i \right)^c = \bigcup_{i=1}^{n} E_i^c$$

Probability Measure

Definition

A *probability measure* on Ω is a function P from subsets of Ω to $[0, 1]$ that satisfies the following rules:

- $P(\Omega) = 1$
- If $A \subset \Omega$, then $P(A) \geq 0$
- If $A_1, A_2, \cdots, A_n, \cdots$ are disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Note: The above rules do not completely determine an assignment of probabililtys to events. They serve only to rule out assignments incositent with our intuitive notions of prob

Probability Measure-cont'd

Some Properties of Probability Measure P

Given any two events $A, B \subset \Omega$

- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$ (i.e., probaility that there is no outcome is 0)
- $P(A) \leq 1$
- If $A \subset B$, then $P(A) \leq P(B)$
- *Addition Law*:

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

Note: The addtion law serves as the general formula to compute the prob of an event of interest which is expressible as a union of events via decomposing the event of interest into "smaller" events

In an undergraduate module, 60% of all students have statistics background, 80% have calculus background, & 50% of all students have both. If a student is selected at random, what is the prob that s/he has background in ❶ at least 1 subject & ❷ exactly 1 subject?

Solution: Let $\begin{cases} A = \{\text{a selected student has statistics background}\} \\ B = \{\text{a selected student has calculus background}\} \end{cases}$

We are given: $P(A) = .6$, $P(B) = .8$, & $P(A \cap B) = .5$

$P(\text{has background in at least 1 subject})$

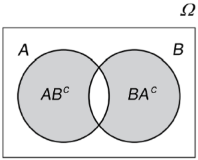
$= P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= .6 + .8 - .5 = .9$

$P(\text{has background in exactly 1 subject})$

$= P(AB^c) + P(BA^c) = P(A \cup B) - P(A \cap B)$

$= .9 - .5 = .4$



Calculating Probability

Depend on the nature of the sample space Ω: *finite* versus *infinite*

When the sample space is finite (i.e., $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$):

- Let
$$P(\{\omega_i\}) = p_i, \qquad i = 1, 2, \cdots, n$$
where $n \geq 1$, called the *cardinality of Ω*, is a finite positive interger denoting the total # of outcomes in Ω
- Computing $P(A)$ is straightforward due to rule of probability:
$$P(A) = P\left(\bigcup_{\omega_i \in A} \{\omega_i\}\right) = \sum_{\omega_i \in A} P(\{\omega_i\}) = \sum_{\{i=1,\cdots,n|\omega_i \in A\}} p_i$$

Simply add probabilities of all the outcomes ω_i in the event A!

Sample Spaces With Equally Likely Outcomes

Many experiments have outcomes equally likely to occur, e.g., coin toss, dice throw, birthday date of a selected student, ...

For these experiments, calculating probability is much easier

Counting Method

For an experiment satisfying that

- the sample space is finite,
- all the n outcomes are equally likely to occur,

the probability of any event A is

$$P(A) = \frac{\#\text{ of outcomes in } A}{\#\text{ of outcomes in } \Omega}$$

Simply count the total number of outcomes in A and in Ω!

Examples-cont'd

Example: flip 2 coins

$\Omega = \{HH, HT, TH, TT\}$ with cardinality $N = 4$

Let A denote the event that at least 1 head is observed

As $A = \{HH, HT, TH\} = \{HH\} \cup \{HT\} \cup \{TH\}$

$$P(A) = P(\{HH\}) + P(\{HT\}) + P(\{TH\})$$

It remains to find the 3 individual probs

Furthermore, if the coin is *fair* (*i.e.,* equally likely to observe head or tail in any toss), *all the 4 possible outcomes should be equally likely to occur* &

$$P(\{HH\}) = P(\{HT\}) = P(\{TH\}) = P(\{TT\}) = \frac{1}{4}$$

Then,

$$P(A) = \frac{\#\text{ of outcomes in } A}{\#\text{ of outcomes in } \Omega} = \frac{3}{4}$$

Counting Methods

In reality, many experiments may be associated with quite large cardinality n

Counting all the outcomes (i.e., cardinality of Ω) is not an easy task; obtaining the cardinality of event A is also prohibitive

Example: the sample space for 50 coin tosses.

Several counting methods or systematic ways for enumeration will be introduced

- Especially useful in computing probs when the sample space is finite
- Some of the ideas can be borrowed or generalized to experiments with infinite sample spaces

Sometimes, the experiment can be decomposed to a sequence of several "simpler" experiments.

Multiplication Principle & its Extension

- If one experiment has $m > 0$ outcomes and another experiment has $n > 0$ outcomes, then there are $\underline{m \times n}$ possible outcomes for the two experiments
- If there are $p > 2$ experiments, where the first experiment has n_1 possible outcomes, the second n_2, $\cdots$, the pth n_p possible outcomes, then there are a total of $\underline{n_1 \times n_2 \times \cdots \times n_p}$ possible outcomes for the p experiments

Example: A coin toss has 2 outcomes: $\{\text{H, T}\}$. 50 coin tosses has 2^{50} outcomes.

Draw a card from a deck of playing cards: 2 experiments with $m = 4$ & $n = 13$ outcomes (as a card is defined by 4 different suits, ♠, ♥, ♣, ♦, & 13 different face values, A, K, Q, J, 10, 9, ... , 3, 2)
$\Rrightarrow$ $4 \times 13 = 52$ possible outcomes

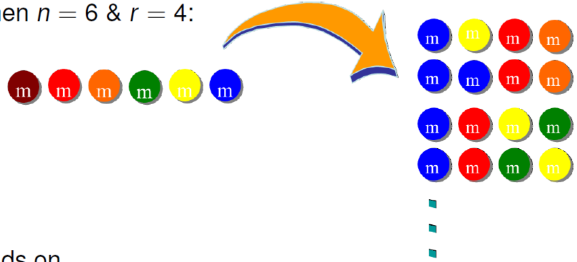
Examine 3 light bulbs: 3 experiments with $n_1 = n_2 = n_3 = 2$ outcomes (💡 or 💡) $\Rrightarrow$ $2 \times 2 \times 2 = 8$ possible outcomes

Singapore Sweep on first Wednesday of every month:
7 experiments with $n_1 = n_2 = \cdots = n_7 = 10$ outcomes
$\Rrightarrow$ $10^7 = 10,000,000$ possible outcomes

Permutations & Combinations

Very often, we would like to address how many ways there are to *select a subset of size r from a group of n* *distinct/distinguishable* *objects* $\{c_1, c_2, \ldots, c_n\}$

e.g., when $n = 6$ & $r = 4$:



It depends on

1. whether we are allowed to *duplicate objects*:
   Sampling without replacement versus *Sampling with replacement*
2. whether the *sequence/order* from which the r objects are selected matters or not

When the ordering matters:

Definition

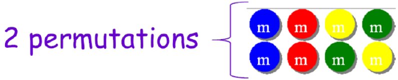
A *permutation* is an ordered arrangement of objects. Selecting a sample of size $r(= 0, 1, \cdots, n)$ from a set of n objects, there are

- n^r permuations under sampling with replacement
- ${}_nP_r = \frac{n!}{(n-r)!} = n(n-1)(n-2)\cdots(n-r+1)$ permuations under sampling without replacement

Recall:

- $n!$ is called *n factorial*
- When n is a positive integer, $n! = n(n-1)(n-2)\cdots 1$. We have the convention $0! = 1$.
- The total number of permutations of n distinct objects is $\frac{n!}{(n-n)!} = n!$

Examples-cont'd



Refer to the picture at Page 29, how many different ordered arrangements of 4 M&M's selected from 6 M&M's of different colors are there?

Solution: Here, $n = 6$ distinct M&M's colors; $r = 4$ selected M&M's colors (subset size). Note that the order of the 4 colors matters and the M&M's cannot be duplicated (i.e, sampling without replacement). The number of permutations is

$$_6P_4 = \frac{6!}{(6-4)!} = \frac{6!}{2!} = 6 \times 5 \times 4 \times 3 = 360$$

There are 10 teaching assistants (TAs) available for grading papers in a test. The test consists of 5 questions, & the Professor wishes to select a different TA to grade each question (at most 1 question per assistant). In how many ways can the TAs be chosen for grading?

Solution: Here, this question is interested in finding the # of permutations with $n = 10$ *distinct* TAs (group size) & $r = 5$ selected TAs (subset size). Note that the order of the 5 selected TAs *matters* (as the 5 questions are *different*) & each TA cannot grade > 1 question (*i.e.*, *sampling without replacement*). The # of permutations is

$$\frac{10!}{(10-5)!} = \frac{10!}{5!} = 10(9)(8)(7)(6) = 30,240$$

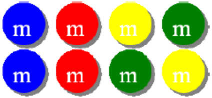
There are n rewards to be distributed randomly by a teacher to n students so that each student gets 1 reward. Suppose that 1 of the rewards is the top prize. The students will queue up to receive a reward from the teacher. Shall a student queue first so as to increase his/her chance of getting the top prize?

Solution: First of all, all possible assignments of n rewards are equally likely to happen and there are $n!$ # of ways/permutations to distribute the rewards

Assume that a student queues at the i-th ($i = 1, \ldots, n$) position in the queue. Imagine that the n rewards also line up and will be distributed one-by-one accordingly. Now, for this student to have the top prize, the i-th reward should be at the i-th position. There are $(n-1)!$ # of ways to distribute the other $n-1$ rewards into the other $n-1$ positions. Hence, the required prob is

$$\frac{(n-1)!}{n!} = \frac{1}{n}$$

- Sometimes, we may be no longer interested in how the objects are arranged, but in the constituents of the subset. For instance, we do not care about the ordering of M& M's colors.



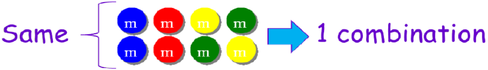
When the ordering does not matter:

Definition

A *combination* is an unordered arrangement/collection of objects. For a set of n distinct objects and a subset of size r, there are ${}_nC_r = \binom{n}{r} = \frac{n!}{(n-r)!r!}$ different combinations under sampling without replacement when $r = 0, 1, \cdots, n$.

Note: $\binom{n}{r}$ read as "n choose r".

Refer to the M& M example on Page 29, how many different combinations of 4 M& M's selected fro m 6 M& M's of different colors are there?



Solution: Here, $n = 6$ distinct M& M's (group size), $r = 4$ selected M& M's (subset size). The order of the 4 colors does not matter and we cannot duplicate the M& M's (i.e., sampling without replacement). The number of combinations is

$$6 choose 4 = \frac{6!}{(6-4)!4!} = \frac{6!}{2!4!} = 15,$$

which is $\ll 360$ (the number of permutations obtained at Page 31)

Binomial Coefficients & The Binomial Theorem

$\binom{n}{r}$ is also called the *binomial coefficient* as it occurs in the expansion

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}.$$

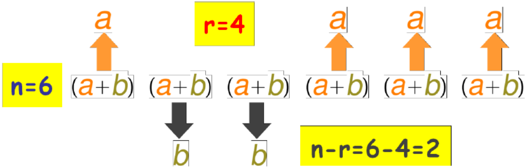
Remark: when $a = 1$ and $b = 1$, there is $2^n = \sum_{r=0}^{n} \binom{n}{r}$.

Example:

$$\begin{aligned}(a+b)^3 &= \binom{3}{0} a^0 b^3 + \binom{3}{1} ab^2 + \binom{3}{2} a^2 b + \binom{3}{3} a^3 b^0 \\ &= b^3 + 3ab^2 + 3a^2 b + a^3\end{aligned}$$

$$(a+b)^n = (a+b)(a+b)\cdots(a+b)$$

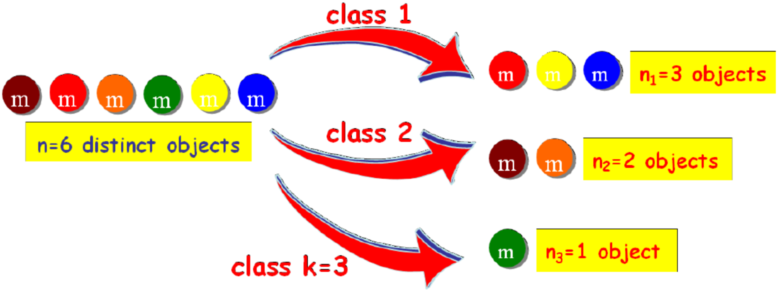
- Relate the binomial expansion to the original selection process:
  - each of the n brackets gives either a or b in the expansion $\Leftrightarrow$ n objects into 2 distinct classes, selected and unselected, respectively
  - $a^r b^{n-r}$ $\Leftrightarrow$ exactly r objects are selected
  - number of terms $a^r b^{n-r}$ $\Leftrightarrow$ number of ways to have exactly r objects selected $\Leftrightarrow$ $\binom{n}{r}$.
- For example, when $n = 6$ and $r = 4$, the coefficient for a^4b^2 is $\binom{6}{4}$. One possibility is



Multinomial Coefficients

Extend the thought about combination from 2 classes to $k \geq 3$ classes: Classes $1, 2, \cdots, k$ with $n_1, n_2, \cdots, n_k$ objects

Example: 6 M& M's, $k = 3$ groups, $n_1 = 3$, $n_2 = 2$ and $n_3 = 1$:



Multinomial Coefficients & The Multinomial Theorem

The number of ways to assign n distinct objects into k distinct classes with n_i objects in the i-th class, $i = 1, 2, \cdots, k$, $\sum_{i=1}^{k} n_i = n$, is

$$\binom{n}{n_1 n_2 \cdots n_k} = \frac{n!}{n_1! n_2! \cdots n_k!},$$

which is called the *multinomial coefficient* as it occurs in the expansion

$$(x_1 + x_2 + \cdots + x_k)^n = \sum \binom{n}{n_1 n_2 \cdots n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k},$$

where the sum is over all nonnegative integers $n_1, n_2, \cdots, n_k$ such that $n_1 + n_2 + \cdots + n_k = n$.

Remark: The assignment/sampling is without replacement as each object is classified to exactly 1 class

NUS National University of Singapore

1. How many different ways of giving 2 M&M's each to 2 kids from 6 M&M's of different colors are there?

   Solution: Here, $n = 6$ *distinct* M&M's/colors (group size), $k = 3$ *distinct* classes (namely, kid 1, kid 2, unassigned), & $n_1 = n_2 = 2$ selected M&M's/colors (class sizes). It implies that $n_3 = 6 - 2 - 2 = 2$. The orders of the colors within classes do *not* matter. The # of ways is

$$\binom{6}{2 \quad 2 \quad 2} = \frac{6!}{2!\,2!\,2!} = 90$$

2. What is the coefficient of $x^2y^2z^3$ in the expansion of $(w + x + y + z)^7$?

   Solution: Take note that $x^2y^2z^3 = w^0x^2y^2z^3$. Here, $n = 7$ & $k = 4$, with $n_1 = 0$, $n_2 = n_3 = 2$ & $n_4 = 3$ (s.t. $n_1 + \cdots + n_4 = n$), the term $x^2y^2z^3$ *exists* & its coefficient is $\dfrac{7!}{0!\,2!\,2!\,3!} = 210$

A Motivating Example: Pokemon Go

- We hope to catch high CP pokemon, e.g., Vulpix
- In some area, the probability for Vulpix to appear is, say, .1.
- However, in case it was known that the user is at level 1, would the chance of him/her finding Vulpix still be .1?
- In case it was known that the user is at level 15, would the chance of him/her finding Vulpix still be .1?

- Of course, it is **intuitive** that the chance for level 1 user to find Vulpix is much smaller!
- The level of Pokemon a user would find is related to his/her level.
- Once it is known that the user has high level, it changes our belief in the probability of getting Vulpix. Indeed, it changes the experiment.
- Given an experiment with all known characteristics and known probs of outcomes, suppose now it happens that some additional information about the experiment is available. It may affect the sample space (i.e., the possible outcomes), and the probs associated with each outcome

Definition

Let $A\&B$ be two events with $P(B) > 0$. The *conditional probability of A given B* is defined to be

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \in [0, 1].$$

- "Given B": an event B has already occurred
- *Reduced sample space* $\Leftrightarrow$ Only outcomes in B, but not in $B^c \subset \Omega$ are possible to occur
  - the sample space for this "new" experiment becomes B rather than Ω & has a probably smaller cardinality
  - it is possible that it is easy to understand A once B has occurred as B contains fewer outcomes than Ω

- There need not be a causal or temporal relationship between A and B.
  Example: The conditional probability that a selected person has height $\geq$ 170cms given that this person weighting $\geq$ 120 lbs?
  Weight and height are related, but larger weight does not cause larger height.
- $P(A|B)$ may or may not be equal to $P(A)$
- In general, $P(A|B)$ (the conditional probability of A given B) is not equal to $P(B|A)$.
- Conditional probabilities can be correctly reversed using *Bayes' rule*.

Suppose that of all individuals buying a certain digital camera, on the spot, 60% also buy a memory card, 40% buy an extra battery, & 30% buy both. Consider randomly selecting a buyer, & let A be the event that a memory card is purchased & B be the event that an extra battery is purchased. Then, we have $P(A) = .6$, $P(B) = .4$, $P(A \cap B) = .3$. Given that the selected buyer purchased an extra battery, prob that a memory card was also purchased is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.3}{.4} = .75$$

$$P(A^c|B) = \frac{P(A^c \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)} = \frac{.1}{.4} = .25$$

Note: $P(A|B) \neq P(A)$ & $\underline{P(A|B) + P(A^c|B) = 1}$

A bin contains 25 light bulbs, of which 5 are good & function at least 30 days, 10 are partially defective & will fail in the second day of use, while the rest are totally defective & won't light up at all. Given that a randomly chosen bulb initially lights up, what is the prob that it will still be working after one week?

Solution: Let G be event that the randomly chosen bulb is in good condition, & T be the event that the randomly chosen bulb is totally defective. This implies that T^c represents the event that the selected bulb is either in good condition or partially defective (*i.e.*, the selected bulb lights up initially). The required conditional prob is

$$P(G|T^c) = \frac{P(GT^c)}{P(T^c)} = \frac{P(G)}{P(T^c)} = \frac{5/25}{15/25} = \frac{1}{3}$$

Note: $P(G|T^c) \neq P(G) = 5/25$

A fair coin is flipped twice. All outcomes in

$$\Omega = \{HH, HT, TH, TT\}$$

are equally likely. What is the prob that both flips result in heads given that the first flip does?

Solution: Let $A = \{HH\}$ & $B = \{HH, HT\}$. Then, $P(B) = 2/4 = 1/2$, & $P(A \cap B) = P(\{HH\}) = 1/4$. So,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{1/2} = \frac{1}{2}$$

Alternatively, since B has occurred, the *reduced sample space* is $B = \{HH, HT\}$. Then, $A \cap B = A = \{HH\}$, & hence, $P(A|B) = 1/2$

Multiplication Law

Multiplication Law

Let $A\&B$ be two events with $P(B) > 0$. Then,

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

- When $P(B)$ and $P(A|B)$ are available or can be easily computed, $P(A \cap B)$ can be obtained as a product
- An alternative formula: When $P(A) > 0$ and $P(B|A)$ are available, $P(A \cap B) = P(A)P(B|A)$
- In practice, for any complex event representable as an intersection of 2 events, its prob can be computed in 2 ways

Examples-cont'd

Four individuals have responded to a request by a blood bank for blood donation. None of them has donated before, so their blood types are unknown. Suppose only type  is desired & only 1 of the 4 actually has this type. If the potential donors are selected at random order for typing, what is the prob that at least 3 individuals must be typed to obtain the desired type?

Solution: Making the identification $B =$ {1st type not } & $A =$ {2nd type not }. Notice that $P(B) = 3/4$ & $P(A|B) = 2/3$. The *multiplication law* yields

$$P(\text{at least 3 individuals are typed}) = P(A \cap B)$$
$$= P(B)P(A|B) = \frac{2}{3} \times \frac{3}{4} = .5$$

Note: In this question, $P(A|B)$ is easily determined due to the idea of *reduced sample space*

Law of Total Probability

Definition

A collection of events $B_1, B_2, \cdots, B_n$ is called a *partition* of size n if

- $B_i \cap B_j = \emptyset$
- $\bigcup_{i=1}^{n} B_i = \Omega$

- $B_1, B_2, \cdots, B_n$ are called mutually exclusive & exhaustive events
- The simplest partition of size 2 including any event B is $\{B, B^c\}$

Law of Total Probability

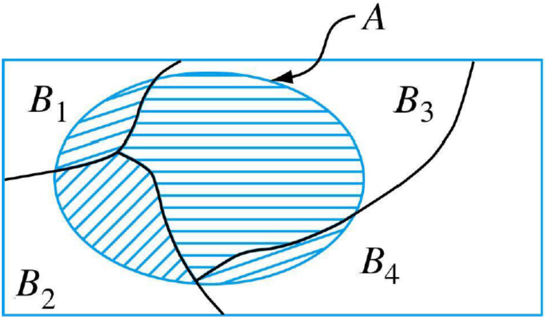
Let $B_1, B_2, \cdots, B_n$ be a partition with $P(B_i) > 0$ for all i. Then, for any event A

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

- To compute $P(A)$ for any A: Choose a partition s.t. all the $2n$ probs at the RHS are available or easily computed

Law of Total Probability-cont'd

- *e.g.*, with a partition of size $n = 4$:



- **The idea**:
  1. break A into 4 (or, in general, n) "smaller" events, namely, $A \cap B_1, \ldots, A \cap B_4$ (or $A \cap B_n$)
  2. compute probs of the 4 (or n) intersections by multiplication law

Of the items produced daily by a factory, 40% come from line 1 & 60% from line 2. Line 1 has a defect rate of 8%, where line 2 has a defect rate of 10%. If an item is chosen at random from the day's production, find the prob that it will not be defective

Solution: Define events: $\begin{cases} D: & \text{item is defective} \\ L_1: & \text{item comes from line 1} \\ L_2: & \text{item comes from line 2} \end{cases}$

From the question, we have $P(L_1) = 1 - P(L_2) = .4$, $P(D^c|L_1) = 1 - P(D|L_1) = .92$, & $P(D^c|L_2) = 1 - P(D|L_2) = .9$. Then,

$$\begin{aligned} P(\text{an item is not defective}) &= P(D^c) \\ &= P(L_1)P(D^c|L_1) + P(L_2)P(D^c|L_2) \\ &= .4(.92) + .6(.9) \\ &= .908 \end{aligned}$$

A chain of video stores sells 3 different brands of DVD players. Of its DVD player sales, 50% are brand 1 & 30% are brand 2. Each manufacturer offers 1–year warranty on parts & labor. It is known that 25% of brand 1's players require warranty repair work, whereas the corresponding percentages for brands 2 & 3 are 20% & 10%, respectively. What is the prob that a randomly selected purchaser has a DVD player that will need repair while under warranty?

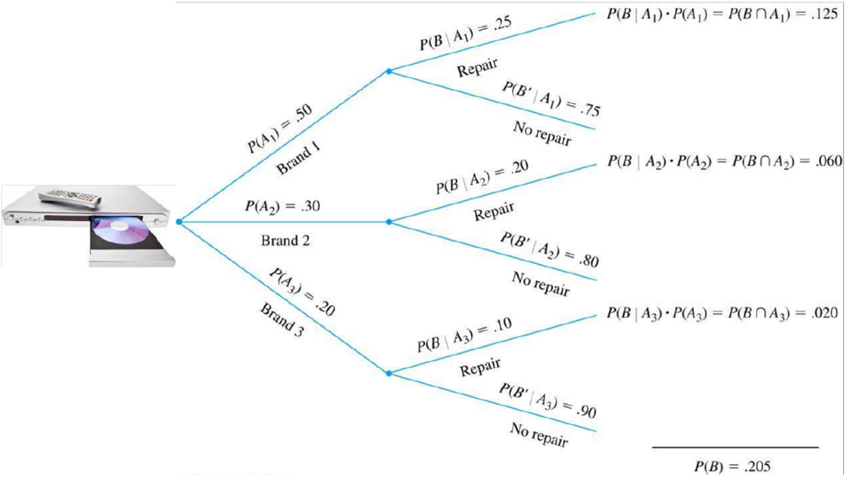
Solution: Let $A_i = \{\text{brand } i \text{ is purchased}\}$, for $i = 1, 2, 3$, & $B = \{\text{needs repair}\}$. Then, we have $P(A_1) = .5$, $P(A_2) = .3$, $P(A_3) = .2$, $P(B|A_1) = .25$, $P(B|A_2) = .2$, $P(B|A_3) = .1$. Treating *$\{A_1, A_2, A_3\}$ as a partition of size 3* & applying the *law of total prob* yield

$$\begin{aligned} P(B) &= P(\{\text{brand 1 \& repair}\} \text{ or } \{\text{brand 2 \& repair}\} \text{ or } \{\text{brand 3 \& repair}\}) \\ &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) \\ &= .5(.25) + .3(.2) + .2(.1) = .205 \end{aligned}$$

Examples-cont'd

- A *tree diagram* is a handy tool for computing probs in *experiments composing of several stages/generations*
- *Components/ingredients* in a tree diagram:
  - *nodes & branches*; total # in different generations depending on the total # of possible outcomes
  - *probs* attached to each branch
- Here, in this example,
  - the initial/1st generation branches correspond to different brands of DVD players $\Rightarrow$ 3 branches in the 1st generation
  - the 2nd generation branches correspond to "needs repair" or "doesn't need repair" $\Rightarrow$ 2 branches at the end of each branch of the 1st generation
  - probs attached to the branches in the 1st generation
  - probs attached to the branches in the 2nd generation: "conditional probs" based on what has happened in the 1st generation

Examples-cont'd



Bayes Rule

Bayes' Rule

Let $B_1, B_2, \cdots, B_n$ be a partition with $P(B_i) > 0$ for all i. Then, for any event A with $P(A) > 0$,

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}, \qquad j = 1, 2, \cdots, n$$

- Numerator: $P(A \cap B_j)$; Denominator: $P(A)$
- Reverse chronological order: usually B_j happens before A in time
  what should have happened before A has occurred?
- As B_j is the event of interest, it provides a hint that one needs to look for a partition containing B_j s.t. the $2n$ probs at the RHS are available or easily computed

Refer to example of DVD players, if a customer returns to the store with a DVD player that needs warranty repair work, what is the prob that it is a brand 1 player? a brand 2 player? a brand 3 player?

Solution: One should apply the Bayes' rule based on the partition $\{A_1, A_2, A_3\}$ to compute these required conditional probs. With $P(B) = .205$ obtained , we have

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{.5(.25)}{.205} = .61$$

$$P(A_2|B) = \frac{P(A_2)P(B|A_2)}{P(B)} = \frac{.3(.2)}{.205} = .29$$

$$P(A_3|B) = \frac{P(A_3)P(B|A_3)}{P(B)} = \frac{.2(.1)}{.205} = .1$$

Note : $P(A_1|B) + P(A_2|B) + P(A_3|B) = 1$

Three telephone lines, A, B & C, are available for calling a cab. The failure rate in connection is 20% for A, 10% for B, & 30% for C. However, line A is more popular & is used for 60% as it is more advertised, whereas line B is used for 30%. A business person failed to connect to a line. What is the prob that s/he used line A?

Solution: Let F be the event of failing to connect to a line. Then,

$$
\begin{aligned}
P(A|F) &= \frac{P(F \cap A)}{P(F)} \\
&= \frac{P(A)P(F|A)}{P(A)P(F|A) + P(B)P(F|B) + P(C)P(F|C)} \\
&= \frac{.6(.2)}{.6(.2) + .3(.1) + .1(.3)} \\
&= .667
\end{aligned}
$$

Independence

- An important concept or notion in prob & stat based on conditional probs
- Recall: The definition of conditional prob enables us to revise the probability $P(A)$ originally assigned to A when we are subsequently informed that another event B has occurred; the "new" prob of A is $P(A|B)$

$$P(A) \stackrel{?}{=} P(A|B)$$

- Intuitively, $P(A|B)$ would be different from $P(A)$ unless knowing B does not tell us anything about A (i.e., "occurrence of B has nothing to do with occurrence of A").

Independence-cont'd

Definition

Events A and B are said to be *independent* if

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

otherwise they are said to be *dependent*.

- In general, multiplication law $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$ is always true for any intersection; the first formula above, $P(A \cap B) = P(A)P(B)$, is a special case following from independence of A & B.
- Independence & disjointness are 2 different concepts
  - conclude disjointness from Venn diagram (no probs involved)
  - independence is defined in terms of probs
  - disjointness means that $P(A|B) = 0 \Rightarrow$ dependence as long as $P(A) \neq 0$ and $P(B) \neq 0$.

Independence of 2 Events

If A & B are independent, then so are $A\&B^c$, $A^c\&B$, and $A^c\&B^c$.

Definition

Three events $A, B\&C$, are said to be *mutually independent* if all the following 4 conditions hold:

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

- A is independent of any event formed by $B\&C$
- Three events are *pairwise independent* if the last 3 conditions hold

1. **Draw a card from a deck of playing cards**: Let $A = \{\text{it is an ace}\}$ & $D = \{\text{it is a } \blacklozenge\}$. Intuitively, knowing the card being an ace should give no information about its suit. Verify this with: $P(A) = 4/52 = 1/13$, $P(D) = 13/52 = 1/4$ &

$$P(A \cap D) = P(\{\text{it is } \blacklozenge A\}) = \frac{1}{52} = \frac{1}{13} \times \frac{1}{4} = P(A)P(D)$$

$\Rrightarrow$ A & D are *indept*

2. **Toss 2 coins**: Suppose that $A = \{\text{1st coin lands a head}\}$, $B = \{\text{2nd coin lands a head}\}$ & $C = \{\text{exactly 1 head is observed}\}$
   - It is clear that $P(B|A) = P(B)$ as whatever happens to the 1st coin does not influence the 2nd coin $\Rrightarrow$ A & B are *indept*
   - Suppose that the coins are *fair* (*i.e.*, $P(A) = P(B) = .5$). Compute $P(C) = P(\{HT, TH\}) = 2/4 = .5$ & thus

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(\{HT\})}{P(\{HT, TH\})} = \frac{.25}{.5} = .5 = P(A)$$

$\Rrightarrow$ A & C are *indept*

Toss 2 fair dice:

1. Let A_6 denote the event that the sum of 2 dice is 6, & B denote the event that the 1st die equals 4

   Then,
   - A_6 = {⚀⚄, ⚁⚃, ⚂⚂, ⚃⚁, ⚄⚀}
   - B = {⚃⚀, ⚃⚁, ⚃⚂, ⚃⚃, ⚃⚄, ⚃⚅}

   So, $A_6 \cap B$ = {⚃⚁}, &

$$\frac{1}{36} = P(A_6 \cap B) \neq P(A_6)P(B) = \frac{5}{36} \times \frac{1}{6}$$

   $\Rrightarrow$ A_6 & B are *dependent*

2. Let A_7 denote the event that the sum of 2 dice is 7. Then,
   $P(A_7) = P($ {⚀⚅, ⚅⚀, ⚁⚄, ⚄⚁, ⚂⚃, ⚃⚂} $) = 6/36 = 1/6$ &
   $P(A_7 \cap B) = P($ {⚃⚂} $) = 1/36$. Hence,

$$\frac{1}{36} = P(A_7 \cap B) = P(A_7)P(B) = \frac{1}{6} \times \frac{1}{6}$$

   $\Rrightarrow$ A_7 & B are *indept*