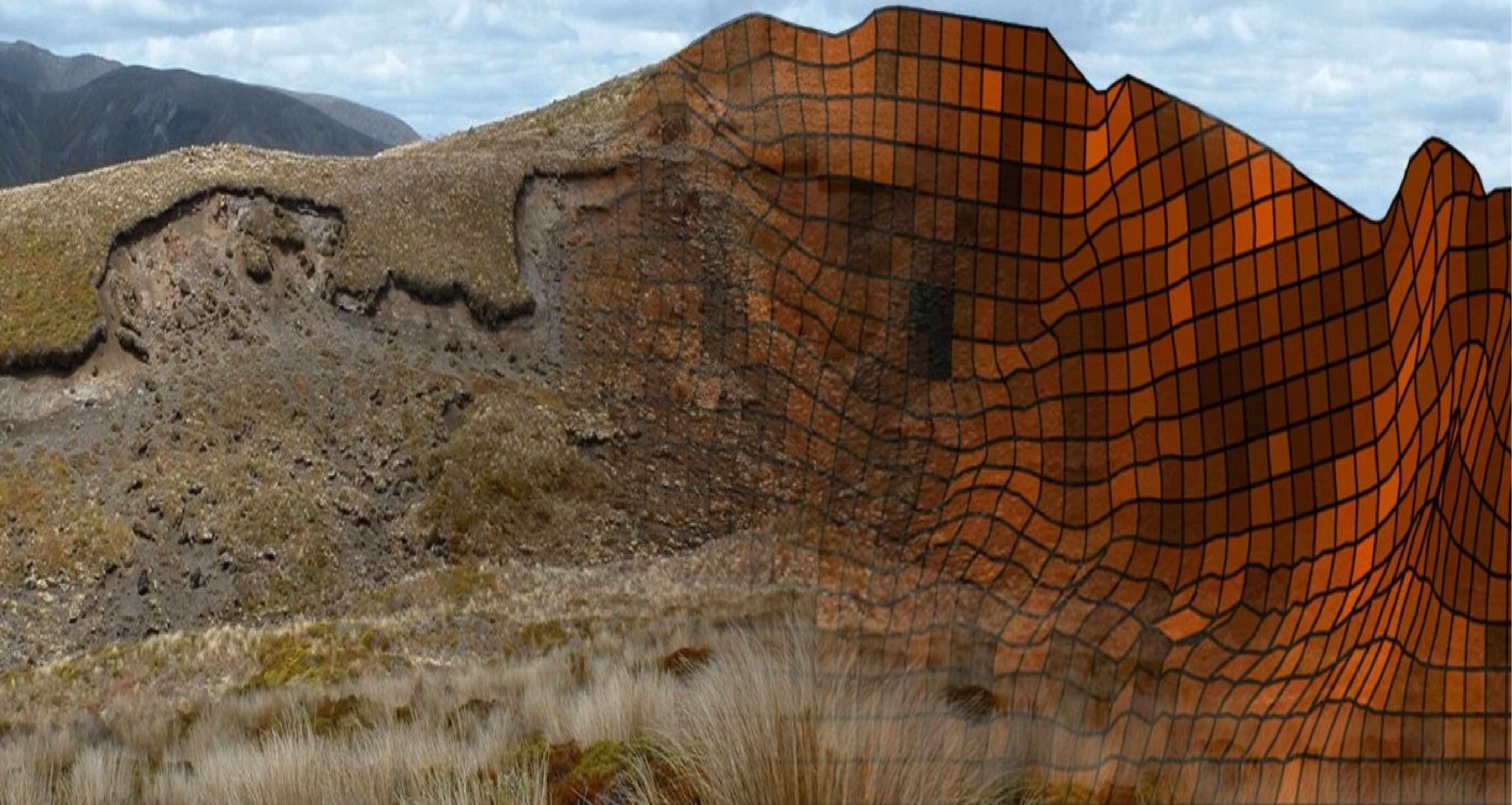
Basic Probability Theory

in *Stochastic Analysis and Inverse Modeling*

Presented by

Gordon A. Fenton



Basic Set Theory

Definitions:

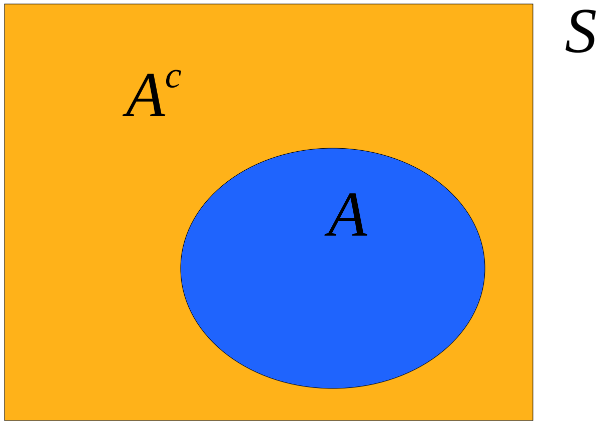
- **Experiment:** Any process that generates a set of data. For example the volume of water passing through an earth dam in a unit of time.
- **Sample space:** The set of all possible outcomes of an experiment. Denoted by S
- **Sample point:** An outcome of a sample space
- **Event:** A subset of sample space. Denoted by A, B, etc.. For example, we might define A to be the event that the flow rate through the dam is less than 0.01 m^3/s
- **Null set:** The empty set is used to denote impossible events. Denoted by Ø. For example the event that a flow rate through an earth dam is both less than 1 m^3/s and greater than 5 m^3/s is impossible so the event is in the null set.

Review of Probability Theory

The relationship between events and the sample space can often be illustrated graphically by means of a Venn Diagram

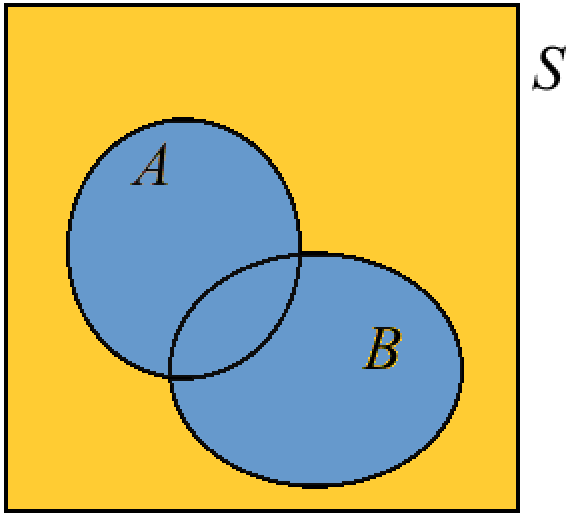
- $A \cup B$ means that either events A or B or both A or B occur
- $A \cap B$ means that both events A and B occur
- A^c means that the event A does not occur
- $A \mid B$ means that the event A occurs given that the event B has occurred.
- S is the sample space (set of all possible outcomes)

Venn Diagram:

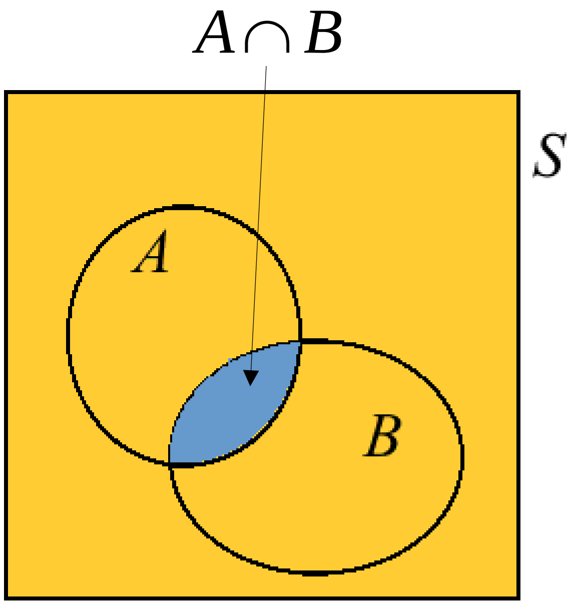


Unions

$A \cup B$



Intersections



Important Results:

$$A \cup A^c = S \qquad A \cap A^c = \varnothing$$

$$A \cap S = A \qquad \varnothing^c = S$$

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

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

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

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

Example: Three piles are being tested statically to failure.

Let A_i denote the event that the i^{th} pile has a capacity exceeding specifications.

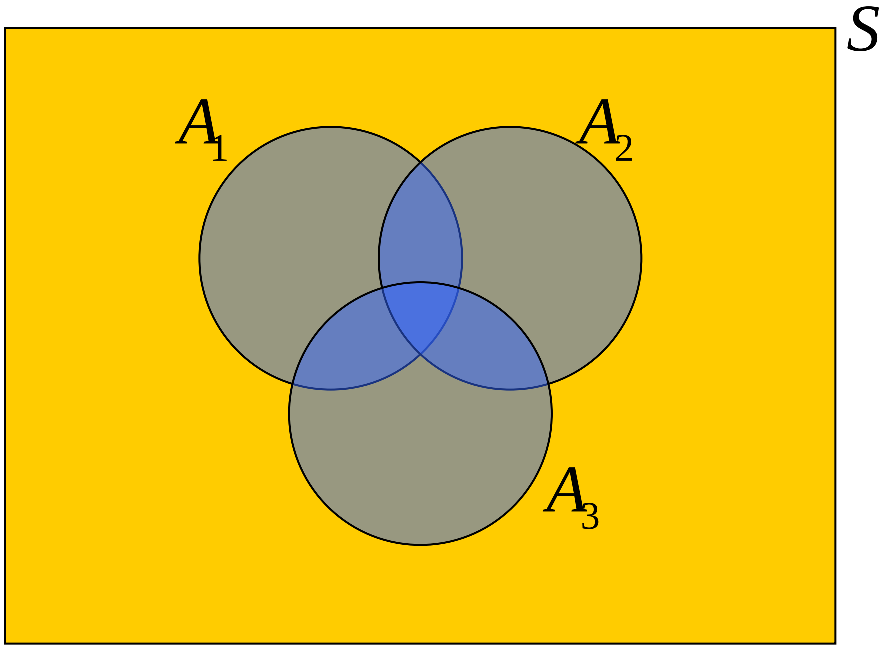
Describe each of the following events using a Venn Diagram and shade the region corresponding to the event.

1. At least one pile has capacity exceeding specifications.
2. All three piles have capacity exceeding specifications.
3. Only the first pile has capacity exceeding specifications.

Solution:

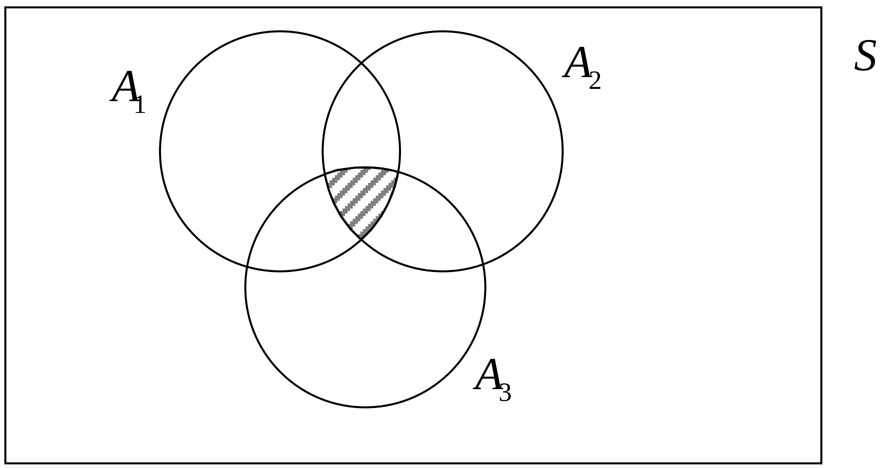
1. At least one pile has capacity exceeding specifications.

$$A_1 \cup A_2 \cup A_3$$



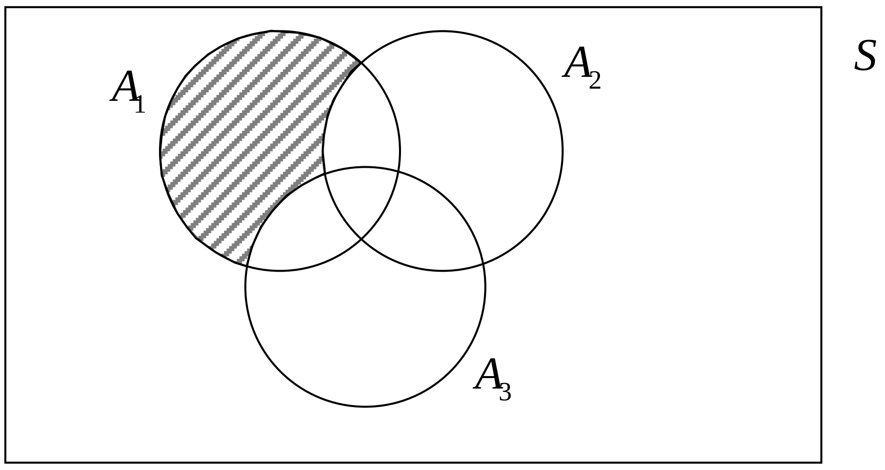
2. All three piles have capacity exceeding specifications.

$A_1 \cap A_2 \cap A_3$



3. Only the first pile has capacity exceeding specifications.

$A_1 \cap A_2^c \cap A_3^c$



Counting Sample Points

If an operation can be performed n_1 ways, and if for each of these, a second operation can be performed n_2 ways, then the two operations can be performed together $n_1 \times n_2$ ways.

Example: A relative density test classifies a soil into five possible states, as follows "very loose“, “loose”, “medium”, “dense” and "very dense".

If 4 soil samples are tested, how many possible ways can the tests proceed if the following conditions are assumed ?

1) The first sample is either "very loose" or "loose", and all four tests are unique (ie, all four tests result in different densities)?
2) The first sample is either "very loose" or "loose", and tests may yield the same results?
3) The first sample is anything but "very loose", and tests may yield the same results?

Solution:

1) The first sample is either "very loose" or "loose", and all four tests are unique (ie, all four tests result in different densities)?

$$2 \times 4 \times 3 \times 2 = 48$$

2) The first sample is either "very loose" or "loose", and tests may yield the same results?

$$2 \times 5 \times 5 \times 5 = 250$$

3) The first sample is anything but "very loose", and tests may yield the same results?

$$4 \times 5 \times 5 \times 5 = 500$$

Permutations and Combinations

The number of **permutations** of r objects selected from n distinct objects *where order counts* is:

$$P_r^n = \frac{n!}{(n-r)!}$$

The number of **combinations** of r objects selected from n distinct objects *without regard to order* is:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Hence there are fewer ways of selecting r objects from n distinct objects if we don't care about the order.

Example: An engineering firm keeps a list of 8 geotechnical engineers. Not all geotechnical engineers are asked to provide a quote on a given request. Determine the number of ways 3 geotechnical engineers can be chosen from the list.

Solution: This is a problem involving **combinations** since when we pick 3 firms we don't care about the order, hence

$$\binom{8}{3} = \frac{8!}{3!5!} = \frac{8\times 7\times 6}{3\times 2\times 1} = 56$$

Example: A company has 7 employees specializing in laboratory testing and 5 employees specializing in field testing. A job requires 2 employees from each area of specialization. In how many ways can the team of 4 be formed?

Solution: This again involves **combinations** since when we pick 2 people from each group we don't care about the order, hence

$$\text{Number of ways we can pick the required team of 4} = \binom{7}{2}\times\binom{5}{2}$$

$$= \frac{7!}{2!5!}\times\frac{5!}{2!3!} = \frac{7\times 6\times 5\times 4}{2\times 2} = 210$$

Event probabilities

The probability of an event A, denoted $\text{P}[A]$, is a number satisfying $0 \le \text{P}[A] \le 1$. We also assume that $\text{P}[\varnothing] = 0$ and P[S]=1

Probability can sometimes be obtained by counting. For example if an experiment can result in any one of N different outcomes, and if exactly m of these outcomes corresponds to event A, then the probability of event A is $\text{P}[A] = m/N$ (assuming equilikely outcomes)

Example: Sixty soil samples have been taken at a site, 5 of which were taken of a liquefiable soil. If 2 of the samples are selected at random from the 60 samples, what is the probability that neither sample will be of the liquefiable soil? (We may assume that specimens are not returned after sampling.)

Solution: $$\text{P[none liquefiable]} = \frac{55}{60} \times \frac{54}{59} = \frac{99}{118} = 0.839$$

alternative way:

$$\text{P[none liquefiable]} = \frac{\binom{55}{2}}{\binom{60}{2}} = \frac{\dfrac{55!}{53!2!}}{\dfrac{60!}{58!2!}} = \frac{55!2!58!}{53!2!60!} = \frac{55 \times 54}{60 \times 59} = 0.839$$

Example: Sixty soil samples have been taken at a site, 5 of which were taken of a liquefiable soil. If 2 of the samples are selected at random from the 60 samples, what is the probability that at least one will be of the liquefiable soil?

$$P[\text{at least one liquefiable}] = 1 - P[\text{none liquefiable}]$$
$$= 1 - 0.839 = 0.161$$

alternative way:

$$P[\text{at least one liquefiable}] = \frac{\binom{5}{1}\binom{55}{1} + \binom{5}{2}\binom{55}{0}}{\binom{60}{2}}$$

$$= \frac{\frac{5!}{1!4!} \times \frac{55!}{1!54!} + \frac{5!}{2!3!} \times \frac{55!}{0!55!}}{\frac{60!}{2!58!}}$$

$$= 0.161$$

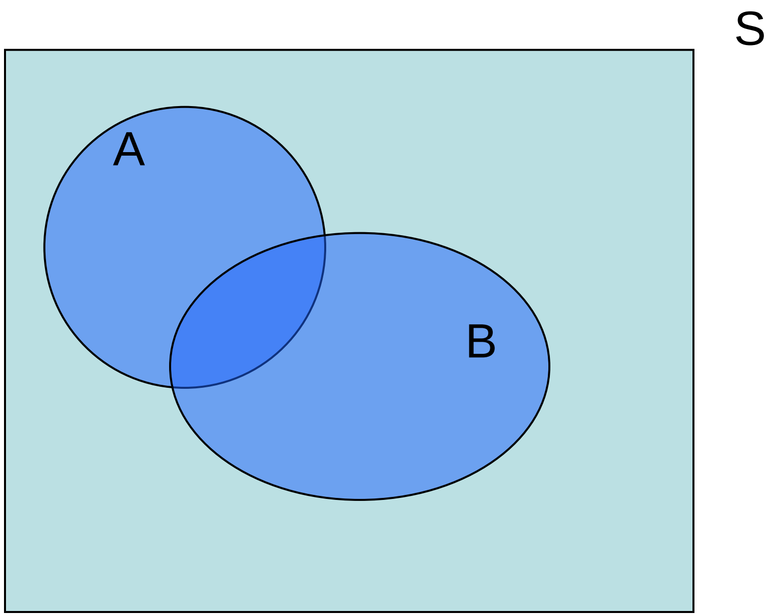
Additive Rules

Sometimes we need to compute the probability of an event expressed in terms of other events.

For example, if A is the event that the company A requests your services and B is the event that company B requests your services, then the event that at least one of the two companies request your services is $A \cup B$

The probability of this is given by the following relationship :

$$P[\, A \cup B \,] = P[\, A \,] + P[\, B \,] - P[\, A \cap B \,]$$



Example: It has been determined that in a particular region, 15% of CPT tests encounter soft clay, 12% encounter gravel and 8% encounter both. If a sounding is selected at random, compute the following probabilities,

1. sounding encounters both soft clay and gravel
2. sounding encounters at least one of these two conditions
3. sounding encounters neither of these two conditions
4. sounding does not encountered gravel
5. sounding encounters a gravel but not a soft clay

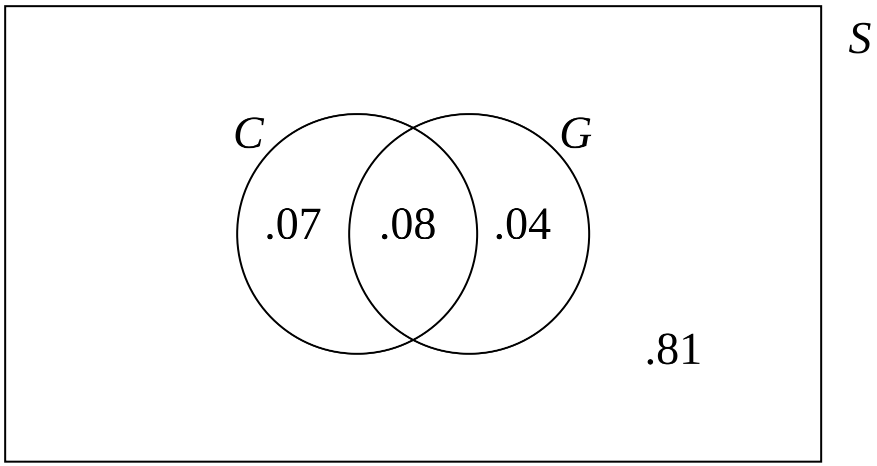
Solution:

Let C be the event that the sounding encounters soft clay

Let G be the event that the sounding encounters gravel

We are given $\mathrm{P}[C] = 0.15, \;\; \mathrm{P}[G] = 0.12, \;\; \mathrm{P}[C \cap G] = 0.08$

A Venn diagram helps to solve this problem



both soft clay and gravel 1) $P[C \cap G] = 0.08$

at least one of these two conditions 2) $P[C \cup G] = P[C] + P[G] - P[C \cap G]$

$$= 0.15 + 0.12 - 0.08$$

$$= 0.19$$

neither of these two conditions 3) $P[C^c \cap G^c] = P[(C \cup G)^c]$

$$= 1 - P[C \cup G]$$

$$= 1 - 0.19$$

$$= 0.81$$

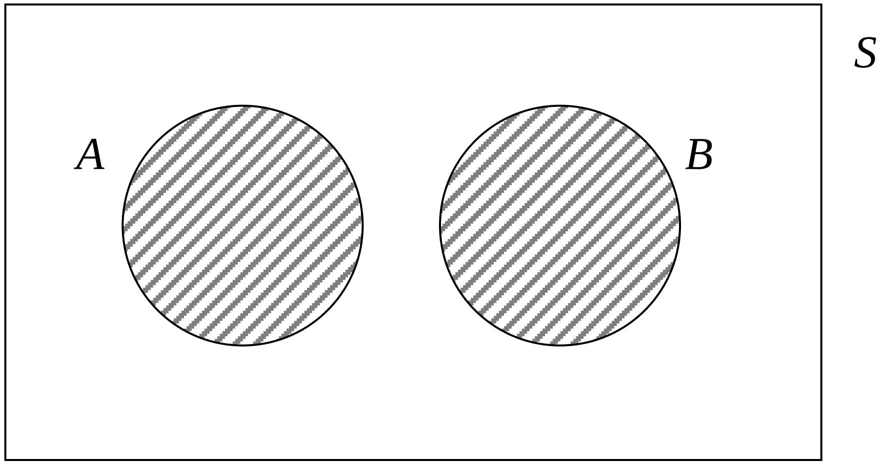
not encountered gravel 4) $P[G^c] = 1 - P[G] = 1 - 0.12 = 0.88$

gravel but not a soft clay 5) $P[G \cap C^c] = 0.04$

Disjoint or Mutually Exclusive Events

Note that if A and B are *disjoint* or *mutually exclusive*, there is no intersection and if one happens the other cannot happen

$$\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B]$$



More generally, if $A_1, A_2, \ldots, A_n$ are *mutually exclusive*, then

$$\mathrm{P}[A_1 \cup A_2 \cup \ldots \cup A_n] = \mathrm{P}[A_1] + \mathrm{P}[A_2] + \ldots + \mathrm{P}[A_n]$$

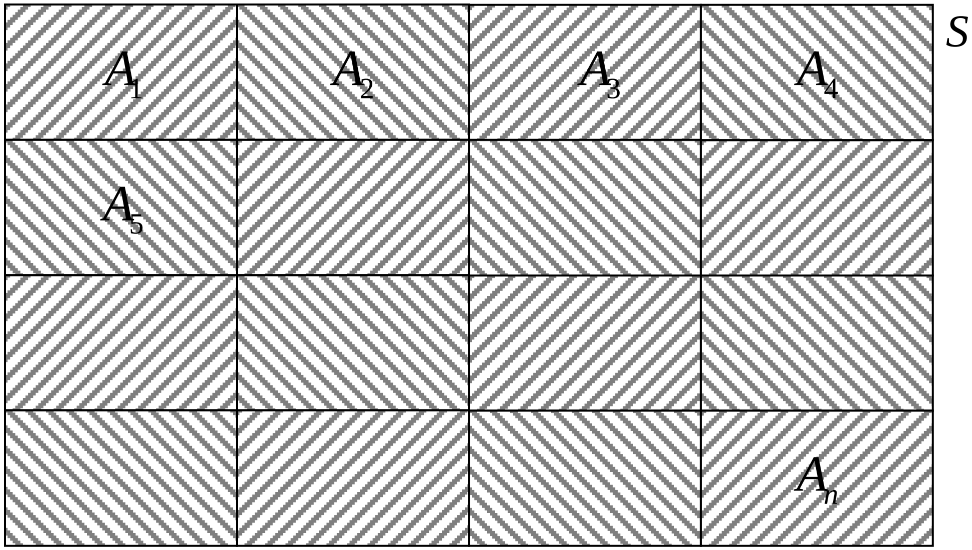
If $A_1, A_2, \ldots, A_n$ are mutually exclusive and entirely occupy the sample space they are said to be collectively exhaustive.

In other words, one of the events must occur.

$$A_1 \cup A_2 \cup \ldots \cup A_n = S$$

and

$$P[A_1 \cup A_2 \cup \ldots \cup A_n] = P[S] = 1$$



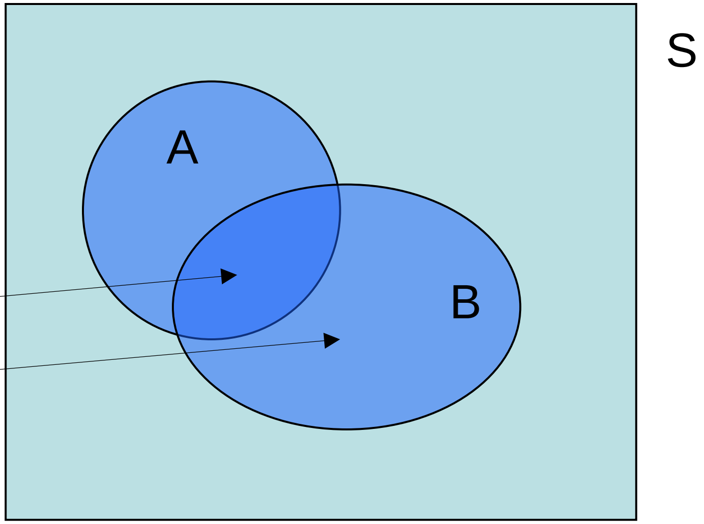
Note: A_i and A_i^c are mutually exclusive and collectively exhaustive so that

$$P[A_i] + P[A_i^c] = 1$$

Conditional Probability

The probability of an event is often affected by the occurrence of other events and/or the knowledge of information relevant to the event. Given two events, A and B, of an experiment, $P[A|B]$ is called the *conditional probability* of A given that B has already occurred.

It is defined by $P[A|B] = \dfrac{P[A \cap B]}{P[B]}$



$$P[A|B] = \frac{\text{area of } A \text{ in } B}{\text{area of } B} = \frac{P[A \cap B]}{P[B]}$$

or $P[A \cap B] = P[A|B]\,P[B]$

Since $P[A \cap B] = P[B \cap A]$ we have

$$P[A|B]\,P[B] = P[B|A]\,P[A] \quad \text{(Bayes' Theorem)}$$

Independence

Two events A and B are **independent** if the occurrence of one event has no influence at all on the probability of occurrence of the other.

This implies that $\mathrm{P}[A \mid B] = \mathrm{P}[A]$, hence $\mathrm{P}[A \cap B] = \mathrm{P}[A] P[B]$

For **disjoint** or **mutually exclusive** events, $\mathrm{P}[A \mid B] = 0$ and $\mathrm{P}[A \cap B] = 0$

Independent events are not mutually exclusive

and mutually exclusive events are not independent!

Example: Four retaining walls, A, B, C, and D, are constructed independently. If their probabilities of sliding failure are estimated to be $\mathrm{P}[A] = 0.01$, $\mathrm{P}[B] = 0.008$, $\mathrm{P}[C] = 0.005$, $\mathrm{P}[D] = 0.015$ respectively, what is the probability that none of them will slide by failing?

Solution: The four events being independent implies that they do not affect the probabilities of the other events occurring.

We want: $\mathrm{P}[A^c \cap B^c \cap C^c \cap D^c]$

$$= \mathrm{P}[A^c]\ \mathrm{P}[B^c]\ \mathrm{P}[C^c]\ \mathrm{P}[D^c] \quad \textit{(since all event are independent)}$$

$$= (1 - \mathrm{P}[A])(1 - \mathrm{P}[B])(1 - \mathrm{P}[C])(1 - \mathrm{P}[D])$$

$$= (1 - 0.01)(1 - 0.008)(1 - 0.005)(1 - 0.015)$$

$$= 0.9625$$

Example: A site consists of 60% sand and 40% silt in separate layers and pockets. At this site, 10% of sand samples and 5% of silt samples are contaminated with trace amounts of arsenic. If a soil sample is selected at random, what is the probability that it is a contaminated sand sample?

Solution: Let A be the event the sample is sand

Let B be the event the sample is silt

Let C be the event the sample is contaminated

Hence $\mathrm{P}[A] = 0.6$, $\mathrm{P}[B] = 0.4$, $\mathrm{P}[C \mid A] = 0.1$, $\mathrm{P}[C \mid B] = 0.05$,

We want to find $\mathrm{P}[A \cap C]$

$$\begin{aligned} \mathrm{P}[A \cap C] &= \mathrm{P}[C \mid A]\mathrm{P}[A] \\ &= 0.1 \times 0.6 = 0.06 \end{aligned}$$

Total Probability Theorem

Sometimes we know the probability of an event in terms of the occurrence of other events and want to compute the *unconditional* probability of the event. For example, when we want to compute the *total* probability of failure of a bridge, we can start by computing a series of simpler problems such as

1. the probability of bridge failure given a maximum static load,
2. the probability of bridge failure given a maximum dynamic traffic load,
3. the probability of bridge failure given an earthquake,
4. the probability of bridge failure given a flood, etc.

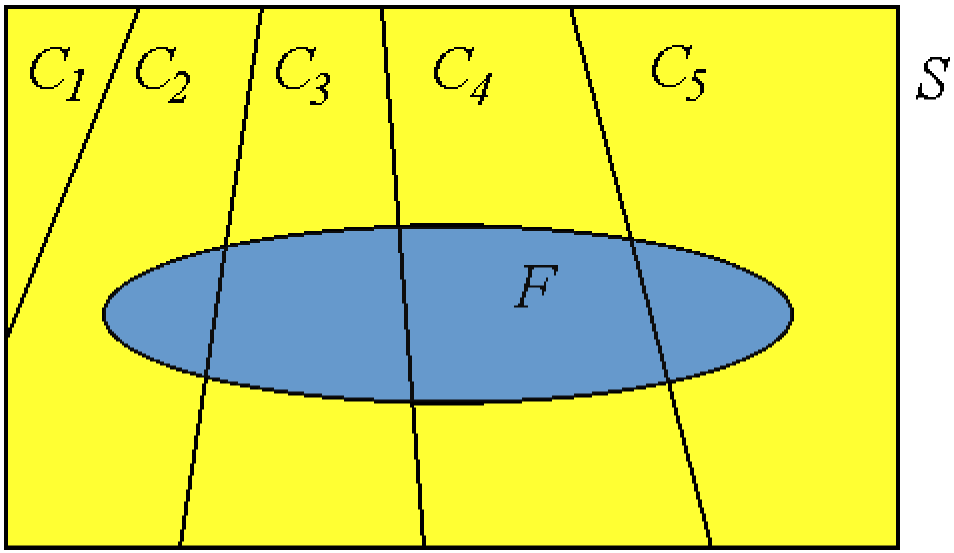
The **Total Probability Theorem** can be used to combine the above probabilities into the unconditional probability of bridge failure. We need to know the above conditional probabilities along with the probabilities that the “conditions” occur (e.g. the probability that the maximum static load will occur during the design life, etc.).

$$P[F] = P\left[(F \cap C_1) \cup (F \cap C_2) \cup (F \cap C_3) \cup (F \cap C_4) \cup (F \cap C_5)\right]$$

$$= P[F \cap C_1] + P[F \cap C_2] + \cdots + P[F \cap C_5] \text{ (since disjoint)}$$

$$= P[F \mid C_1]P[C_1] + P[F \mid C_2]P[C_2] + \cdots + P[F \mid C_5]P[C_5]$$

Total Probability Theorem



$P[F \cap C_1]$

$P[F \cap C_2]$

$P[F \cap C_3]$

$P[F \cap C_4]$

$P[F \cap C_5]$

are *disjoint*

C_1 = event that no extreme effects occur

C_2 = event of maximum static load

C_3 = event of maximum dynamic load

C_4 = event of an earthquake

C_5 = event of a flood

Bayes' Theorem

Since $$\mathrm{P}[A \cap E] = \mathrm{P}[E \cap A]$$

we must have $$\mathrm{P}[A \mid E]\mathrm{P}[E] = \mathrm{P}[E \mid A]\mathrm{P}[A]$$

which leads to Bayes' Theorem:

$$\mathrm{P}[A \mid E] = \frac{\mathrm{P}[E \mid A]\mathrm{P}[A]}{\mathrm{P}[E]}$$

More on Bayes' Theorem

a posteriori → $P[A|E]$; *a priori* → $P[A]$

$$P[A|E] = \frac{P[E|A]\,P[A]}{P[E]}$$

Bayes' theorem can be used to update probabilities as additional information becomes available.

Example: A contaminated soil has one of three possibilities:

A. only toxin 1 present (state *A*) → remediation scheme 1
B. only toxin 2 present (state *B*) → remediation scheme 2
C. both toxins present (state *C*) → remediation scheme 3

If state C is true, assume that toxins 1 and 2 occur in equal quantities.

Assume that we have no prior bias about which state is true.
Thus, choose our "*a priori*" values to be

$$P[A] = 1/3$$
$$P[B] = 1/3$$
$$P[C] = 1/3$$

A sample is taken which yields trace amounts of toxin 1.

Let this event be E_1

What is the updated probability of states A, B and C given that event E_1 has occured?

If A is true (only toxin 1), then the probability of observing trace amounts of toxin 1 given A is 1.0 → $\mathrm{P}\left[E_1 \middle| A\right] = 1$

If B is true (only toxin 2) then → $\mathrm{P}\left[E_1 \middle| B\right] = 0$

If C is true (both toxins) then → $\mathrm{P}\left[E_1 \middle| C\right] = 1/2$
(assume toxins occur in equal proportions if C is true)

Probability of event E_1 from Total Probability Theorem

$$\mathrm{P}\left[E_1\right] = \mathrm{P}\left[E_1 \middle| A\right]\mathrm{P}\left[A\right] + \mathrm{P}\left[E_1 \middle| B\right]\mathrm{P}\left[B\right] + \mathrm{P}\left[E_1 \middle| C\right]\mathrm{P}\left[C\right]$$
$$= (1)(1/3) + (0)(1/3) + (1/2)(1/3)$$
$$= 1/2$$

Updated probabilities, given that Toxin 1 has been found in the first sample

Compute the *a posteriori* values

$$\mathrm{P}[A \mid E_1] = \frac{\mathrm{P}[E_1 \mid A]\mathrm{P}[A]}{\mathrm{P}[E_1]} = \frac{(1)(1/3)}{(1/2)} = 2/3$$

$$\mathrm{P}[B \mid E_1] = \frac{\mathrm{P}[E_1 \mid B]\mathrm{P}[B]}{\mathrm{P}[E_1]} = \frac{(0)(1/3)}{(1/2)} = 0$$

$$\mathrm{P}[C \mid E_1] = \frac{\mathrm{P}[E_1 \mid C]\mathrm{P}[C]}{\mathrm{P}[E_1]} = \frac{(1/2)(1/3)}{(1/2)} = 1/3$$

These will be the *a priori* values for the next update.

A second sample is taken which again yields trace amounts of toxin 1.

Let this event be E_2 and update once more.

If A is true (only toxin 1), then the probability of observing trace amounts of toxin 1 given A is 1.0 → $\mathrm{P}\left[E_2 \middle| A\right] = 1$

If B is true (only toxin 2) then → $\mathrm{P}\left[E_2 \middle| B\right] = 0$

If C is true (both toxins) then → $\mathrm{P}\left[E_2 \middle| C\right] = 1/2$
(assume toxins occur in equal proportions if C is true)

Probability of event E_2 from Total Probability Theorem

$$\begin{aligned}\mathrm{P}\left[E_2\right] &= \mathrm{P}\left[E_2 \middle| A\right]\mathrm{P}\left[A\right] + \mathrm{P}\left[E_2 \middle| B\right]\mathrm{P}\left[B\right] + \mathrm{P}\left[E_2 \middle| C\right]\mathrm{P}\left[C\right] \\ &= (1)(2/3) + (0)(0) + (1/2)(1/3) \\ &= 5/6\end{aligned}$$

Updated probabilities, given that Toxin 1 has once again been found in the second sample

Compute the *a posteriori* values

$$P[A|E_2] = \frac{P[E_2|A]P[A]}{P[E_2]} = \frac{(1)(2/3)}{(5/6)} = 4/5$$

$$P[B|E_2] = \frac{P[E_2|B]P[B]}{P[E_2]} = \frac{(0)(0)}{(5/6)} = 0$$

$$P[C|E_2] = \frac{P[E_2|C]P[C]}{P[E_2]} = \frac{(1/2)(1/3)}{(5/6)} = 1/5$$

These will be the *a priori* values for the next update and so on….

Now suppose that a 3rd sample is taken and it contains trace amounts of toxin 2. Let this event be E_3 .

We know in this case that we must have State C since we have now encountered both toxins. I.e., we must have

$$\mathrm{P}\left[A \mid E_3\right] = 0$$

$$\mathrm{P}\left[\mathrm{B} \mid E_3\right] = 0$$

$$\mathrm{P}\left[\mathrm{C} \mid E_3\right] = 1$$

Let's see if Bayes' Theorem gives this result.

If A is true (only toxin 1), then the probability of observing trace amounts of toxin 2 given A is 0.0 → $\mathrm{P}\left[E_3 \middle| A\right] = 0$

If B is true (only toxin 2) then → $\mathrm{P}\left[E_3 \middle| B\right] = 1$

If C is true (both toxins) then → $\mathrm{P}\left[E_3 \middle| C\right] = 1/2$
(assume toxins occur in equal proportions if C is true)

Probability of event E_3 from Total Probability Theorem

$$\begin{aligned}\mathrm{P}\left[E_3\right] &= \mathrm{P}\left[E_3 \middle| A\right]\mathrm{P}\left[A\right] + \mathrm{P}\left[E_3 \middle| B\right]\mathrm{P}\left[B\right] + \mathrm{P}\left[E_3 \middle| C\right]\mathrm{P}\left[C\right] \\ &= (0)(4/5) + (1)(0) + (1/2)(1/5) \\ &= 1/10\end{aligned}$$

Updated probabilities, given that Toxin 2 has now been found in the third sample

Compute the *a posteriori* values

$$P[A|E_3] = \frac{P[E_3|A]P[A]}{P[E_3]} = \frac{(0)(4/5)}{(1/10)} = 0$$

$$P[B|E_3] = \frac{P[E_3|B]P[B]}{P[E_3]} = \frac{(1)(0)}{(1/10)} = 0$$

$$P[C|E_3] = \frac{P[E_3|C]P[C]}{P[E_3]} = \frac{(1/2)(1/5)}{(1/10)} = 1$$

as expected. No further testing is now required…