# Applying the Rules of Probability

We work with the following “rules” in order to calculate event probabilities:

**De Morgan**’s rule for complement of a union (applies also for more events):
$\overline{E_1 \cup E_2} = \overline{E_1} \cap \overline{E_2}$

**De Morgan**’s rule for complement of an intersection (applies also for more events):
$\overline{E_1 \cap E_2} = \overline{E_1} \cup \overline{E_2}$

Probability of the **complement**:
$P(\overline{E}) = 1 - P(E)$

**Union** rule:
$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1\,E_2)$

**Inclusion-exclusion** rule (applies also for more events):
$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1\,E_2) - P(E_1\,E_3) - P(E_2\,E_3) + P(E_1\,E_2\,E_3)$

**Conditional** probability rule:
$P(E_1|E_2) = \frac{P(E_1\,E_2)}{P(E_2)}$

**Multiplication** rule:
$P(E_1E_2) = P(E_1|E_2)P(E_2) = P(E_2|E_1)P(E_1)$

**Bayes**’ rule:
$P(E_1|E_2) = \frac{P(E_2|E_1)}{P(E_2)}P(E_1)$

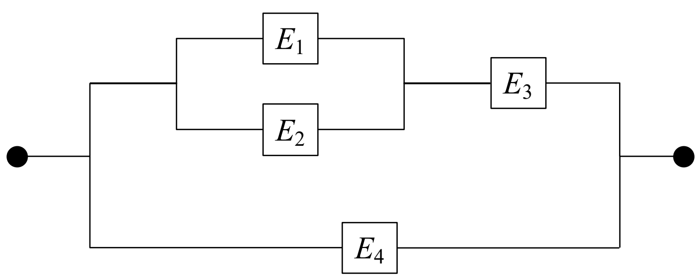
Rule (theorem) of **total probability** (for MECE events):
$P(A) = \sum_{i=1}^{N} P(A\,|\,E_i)\,P(E_i)$

**Statistical independence** is checked by:
$P(E_1|E_2) = P(E_1)$ or $P(E_2|E_1) = P(E_2)$

**Statistical independence** has the consequence:
$P(E_1E_2) = P(E_1)P(E_2)$

# Transportation Network

This problem considers the figure shown below. It is instructive to think of that figure as a transportation network between two locations indicated by solid dots. The rectangles are thought of as infrastructure objects, such as bridges. Following that analogy, $E_1$ denotes the event that Bridge 1 is closed, and similar for $E_2$, $E_3$, and $E_4$. The objective is to determine the probability that transportation between the two locations is either possible or not. Interestingly, this is an introductory example that also appears far down the road, so to speak, in the topic entitled System Reliability. At that time, the figure below will be referred to as a "reliability block diagram" with applications far beyond transportation networks. For now, the simplifying assumption is made that the components of the network, i.e., the events $E_i$ are statistically independent. It is that assumption that makes the problem solvable here, but statistical independence is hard to come by, in realistic problems. For example, for the considered network, it may be the same hazard that lurks behind each failure event; that may introduce statistical dependence. Later, in System Reliability, the components may be structural members or limit-state functions that share parameters, again introducing statistical dependence.




Let the individual event probabilities be:

```
PofEi = 0.2;
```

The probability that the communication between the locations is unavailable is, by inspection of the figure above:

$P(E_1\, E_2\, E_4 \cup E_3\, E_4)$

The union rule gives:

$P(E_1\, E_2\, E_4 \cup E_3\, E_4) = P(E_1\, E_2\, E_4) + P(E_3\, E_4) - P(E_1\, E_2\, E_3\, E_4)$

Statistical independence gives:

$P(E_1\, E_2\, E_4) = P(E_1)P(E_2)P(E_4)$

$P(E_3\, E_4) = P(E_3)P(E_4)$

$P(E_1\, E_2\, E_3\, E_4) = P(E_1)P(E_2)P(E_3)P(E_4)$

That gives the final result:

```
Pclosed = PofEi^3 + PofEi^2 - PofEi^4
```

which yields: 0.0464

The probability that communication between the two cities is open can now be determined in two ways. Either by considering the complement of the previous solution and using de Morgan's rules, or inspecting the figure again. Both give the same result; here, we start with the complement of the previous calculations, and the use of de Morgan's rules:

$$P(\overline{E_1\, E_2\, E_4 \cup E_3\, E_4}) = P(\overline{E_1\, E_2\, E_4} \cap \overline{E_3\, E_4}) = P(\ (\overline{E_1} \cup \overline{E_2} \cup \overline{E_4}) \cap (\overline{E_3} \cup \overline{E_4})\ )$$

However, the two parentheses on either side of the intersection symbol are not statistically independent, because $\overline{E_4}$ appears in both. We must therefore proceed in order to obtain an answer. Caution is required when applying the distributive rule, essentially "multiplying" the two parentheses. To that end, observe that $\overline{E_4}$ appears in both parenthesis. For that reason, we pull that even out and place it in a separate entry of the union of events that now emerge. Then, the distributive rule allows us to "multiply" $\overline{E_3}$ into the parenthesis where now $\overline{E_1}$ and $\overline{E_2}$ remain. The result is:

$$... = P(\overline{E_1}\, \overline{E_3} \cup \overline{E_2}\ \ \overline{E_3} \cup \overline{E_4})$$

One may inspect the figure in order to establish and verify that expression. Regardless, the result is:

```
Popen = (1 - PofEi)^2 + (1 - PofEi)^2 + (1 - PofEi) - 3 (1 - PofEi)^3 + (1 - PofEi)^4
```

which yields: 0.9536

Notice the unit sum of the two answer obtained, which makes sense because one is the complement of the other:

```
Pclosed + Popen
```

which yields: 1.

# Facility Exposed to Hazards

En engineering facility is subjected to a hazard, e.g., ground shaking, with three possible intensities: S=strong, M=moderate, W=weak. The following probabilities are given:

```
PofS = 0.03;
PofM = 0.3;
PofW = 0.67;
```

For each hazard level, there is a probability that the facility will fail. The higher the intensity of the hazard, the higher the conditional failure probability is:

```
PofFgivenS = 0.25;
PofFgivenM = 0.05;
PofFgiveW = 0.015;
```

The probability of failure of the facility in an impending occurrence of the hazard is obtain by the rule of total probability:

```
PofF = PofFgivenS PofS + PofFgivenM PofM + PofFgiveW PofW
```

which yields: 0.03255

Next, let’s say that we receive word that the facility has failed. In that case, the probability that the earthquake was of moderate strength is obtained via Bayes’ rule:

$$\text{PofMgivenF} = \frac{\text{PofFgivenM}}{\text{PofF}}\,\text{PofM}$$

which yields: 0.460829

## Testing for Defect in Wood Products

Consider a rare defect with a certain type of engineered wood product. For each specimen, we define the event D=defect is present and T=test shows defect. During production there is a small probability,

```
PofD = 0.008;
```

that any one product comes out defect. The products are tested by a good but imperfect test device. Specifically, if a product is defect then there is a probability,

```
PofTgivenD = 0.9;
```

that the test device will indicate a flaw. On the other hand, if a product does not have the defect then there is a probability,

```
PofTgivenNotD = 0.07;
```

that the test devise will still categorize it as a defect product. First, let us examine the probability that a product that has tested as defect is actually defect. That probability is available via Bayes’ rule. However, before we can evaluate that formula, we recognize that the denominator in Bayes rule is P(T), i.e., the probability of the test showing a defect, for any given specimen. As often is the case in applications of Bayes' rule, that denominator-probability is evaluated by means of the rule of total probability:

$$\text{PofT} = \text{PofTgivenD PofD} + \text{PofTgivenNotD} \, (1 - \text{PofD})$$

which yields: 0.07664

We are then ready to apply Bayes' rule:

$$\text{PofDgivenT} = \frac{\text{PofTgivenD}}{\text{PofT}} \text{PofD}$$

which yields: 0.0939457

It may appear strange that the probability of a defect is so small even when the indicator says there is a defect present. The reasons for this is the low probability of defects in the general population of wood products, combined with the imperfection in the testing device. This kind of example is sometimes offered to medical professionals, who are making decisions about treatment based on testing outcomes from imperfect devices. Notice that if a second test is conducted, then the result is compounded with the previous calculations simply by reapplying Bayes' rule and the rule of total probability, but now with P(D | T) from earlier serving as P(D). That gives a far greater probability of defect, given a second test result suggesting a flaw:

$$\text{PofT} = \text{PofTgivenD PofDgivenT} + \text{PofTgivenNotD} \, (1 - \text{PofDgivenT});$$

$$\text{PofDgivenT} = \frac{\text{PofTgivenD}}{\text{PofT}} \text{PofDgivenT}$$

which yields: 0.571388

## Markovian Wood Product Defects

This example continues the consideration of defects in wood products, now without explicit test results, but rather with probabilistic information about the dependence of one wood product on those tested immediately prior. This example takes features from Problems 2.20 and 2.21 in the textbook on Probabilistic Concepts in Engineering by Ang & Tang. The word Markovian, used in the heading, means dependence on the previous test result, but not further back. That will be the starting point, but it is extended later in the example. Let the probability of defect in any given product be as given in the previous example. Denote the presence of a defect in three consecutively manufactured products be denoted by $D_1$, $D_2$, and $D_3$. For now, the Markovian assumption stands; i.e., the presence of a defect in any given specimen depends only on the specimen manufactured immediately prior to this one. In fact, if the previous specimen was defect, then that triples the probability of a defect:

```
PofD2givenD1 = 3 PofD;
PofD3givenD2 = 3 PofD;
```

Now, let’s calculate some probabilities with the given information. First, for reference, the probability of a defect for a single specimen, without any information about past specimens is:

```
PofD
```

which yields: 0.008

Next, the probability that two consecutive wood products are defect, is, using the multiplication rule:

```
PofD1andD2 = PofD2givenD1 PofD
```

which yields: 0.000192

Next, we calculate the probability that three consecutive wood products are defect, again using the multiplication rule:

```
PofD1andD2andD3 = PofD3givenD2 PofD2givenD1 PofD
```

which yields: $4.608 \times 10^{-6}$

In order to contrast those calculations with a non-Markovian situation, we now postulate that if the two previous specimens have been defect, then the probability of a defect in the next specimen is not triple P(D) but rather double:

```
PofD3givenD2andD1 = 2 PofD;
```

The probability of three consecutive defect products is now:

$P(D_1\, D_2\, D_3) = P(D_3\,|\,D_1\, D_2)\; P(D_1\, D_2) = \; P(D_3\,|\,D_1\, D_2)\; P(D_2\,|\,D_1)\;\; P(D_1)$, which evaluates to:

```
PofD1andD2andD3 = PofD3givenD2andD1 PofD2givenD1 PofD
```

which yields: $3.072 \times 10^{-6}$

Now back to the Markovian case, if we wish to calculate the probability that at least one out of three products are defect, then we can use the same strategy as employed then calculating the probability of occurrence, i.e., any number of occurrences, in Bernoulli trials or the Poisson process. That is, we calculate the probability of the complement:

P(at least one defect out of three products)

= 1 – P(no defects in three products)
$= 1 - P(\overline{D_1}\,\overline{D_2}\,\overline{D_3})$
$= 1 - P(\overline{D_3} \mid \overline{D_2})\, P(\overline{D_2} \mid \overline{D_1})\, P(\overline{D_1})$ ... Eq. (*)

However, we do not have the probabilities $P(\overline{D_3} \mid \overline{D_2})$ and $P(\overline{D_2} \mid \overline{D_1})$. This problem is neatly solved by Ang & Tang by first giving the generic hint:

$$P(\overline{E_1} \cup \overline{E_2}) = 1 - P(\overline{\overline{E_1} \cup \overline{E_2}}) = 1 - P(E_1\, E_2) \quad \text{... Eq.(**)}$$

In order to proceed, expand the left-hand side in Eq. (**) in the following manner:

$$P(\overline{E_1} \cup \overline{E_2}) = P(\overline{E_1}) + P(\overline{E_2}) - P(\overline{E_1}\,\overline{E_2})$$

where the last term contains what we seek:

$$P(\overline{E_1}\,\overline{E_2}) = P(\overline{E_2} \mid \overline{E_1}) P(\overline{E_1})$$

Next, we spell out the last term in Eq. (**) in a similar manner:

$$P(E_1\, E_2) = P(E_2 \mid E_1)\, P(E_1)$$

Substituting everything back into Eq. (**) yields:

$$P(\overline{E_1}) + P(\overline{E_2}) - P(\overline{E_2} \mid \overline{E_1})\, P(\overline{E_1}) = 1 - P(E_2 \mid E_1)\, P(E_1)$$

Solving for the sought conditional probability yields:

$$P(\overline{E_2} \mid \overline{E_1}) = \frac{P(\overline{E_1}) + P(\overline{E_2}) - (1 - P(E_2 \mid E_1)\; P(E_1))}{P(\overline{E_1})}$$

For the numbers given earlier in this example, that conditional probability is:

$$\text{PnotD3givenNotD2} = \frac{(1 - \text{PofD}) + (1 - \text{PofD}) - (1 - \text{PofD2givenD1}\;\text{PofD})}{(1 - \text{PofD})}$$

which yields: 0.992129

In turn, that gives the following result when evaluating Eq. (*):

PofAtLeasOneDefectInThree = 1 - PnotD3givenNotD2$^2$ (1 - PofD)

which yields: 0.0235545

Notice that if we had rather wished to calculate the probability that at least one out of TWO products are defect, then that case is simpler:

$$P(D_1 \cup D_2) = P(D_1) + P(D_2) - P(D_1 D_2) = P(D_1) + P(D_2) - P(D_2 | D_1) P(D_1)$$

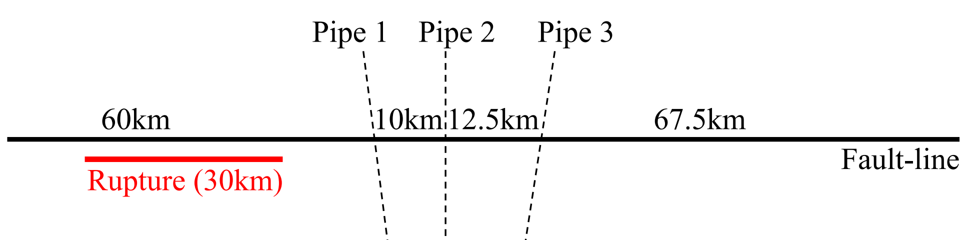
The value of that expression is:

PofAtLeasOneDefectInTwo = PofD + PofD - PofD2givenD1 PofD

which yields: 0.015808

## Pipelines Cut by Fault-line Rupture

Consider three pipelines crossing an earthquake fault-line. A rupture may occur along that fault-line, causing an earthquake. We assume that the impending rupture is exactly 30km long, and that all of the rupture will occur within the confines of the fault-line shown in the figure below. Furthermore, the rupture is equally likely to occur anywhere within the fault-line. If the rupture crosses a pipeline then that pipeline is assumed to fail. This is an example of obtaining probabilities by geometrical considerations, but notice that *conditional* probabilities can *not* be determined in that manner. That point is relevant, because we here wish to determine whether the individual pipe failures are statistically independent.




Notice that the total "outcome space" is

outcomeSpace = (60 + 10 + 12.5 + 67.5) - 30

which yields: 120.

The subtraction of 30km in that expression is because the rupture cannot extend beyond the ends of the fault-line. Also note that a pipe fails if the rupture occurs within a 30km portion of the outcome space. This means that the probability of failure for any individual pipeline is "event" / "outcome space":

$$\text{PofAnyPipe} = \frac{30}{\text{outcomeSpace}}$$

which yields: 0.25

In order to test for statistical independence, we must also evaluate P(failure of pipe 1 | failure of pipe 2), etc. We cannot determine conditional probabilities by geometrical considerations, but the conditional probability rule comes to rescue:

$$P(F_i \mid F_j) = \frac{P(F_i F_j)}{P(F_j)}$$

The probability of joint (simultaneous) pipe failures can indeed be determined by geometrical considerations:

$$\text{PofPipe1andPipe2} = \frac{30 - 10}{\text{outcomeSpace}}$$

which yields: 0.166667

$$\text{PofPipe1andPipe3} = \frac{30 - 10 - 12.5}{\text{outcomeSpace}}$$

which yields: 0.0625

$$\text{PofPipe2andPipe3} = \frac{30 - 12.5}{\text{outcomeSpace}}$$

In turn, that gives the following conditional probabilities:

$$\text{PofPipe1givenPipe2} = \frac{\text{PofPipe1andPipe2}}{\text{PofAnyPipe}}$$

which yields: 0.666667

$$\text{PofPipe1givenPipe3} = \frac{\text{PofPipe1andPipe3}}{\text{PofAnyPipe}}$$

which yields: 0.25

$$\text{PofPipe2givenPipe3} = \frac{\text{PofPipe2andPipe3}}{\text{PofAnyPipe}}$$

Interestingly, pipelines 1 and 3 are spaced exactly by the amount that makes their failure events statistically independent, i.e., $P(F_1 \,|\, F_3) = P(F_3 \,|\, F_1) = P(F_1) = P(F_3) = 0.25$. In other words, if we receive word that pipeline 1 has failed, then pipeline 3 is exactly at such a distance that its probability of failure does not change.