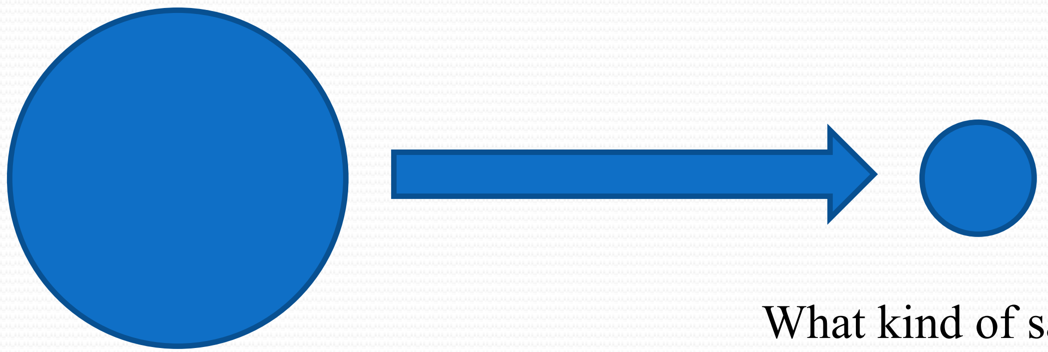
BIM 105
Probability and Statistics for Biomedical Engineers

David M. Rocke

Department of Biomedical Engineering

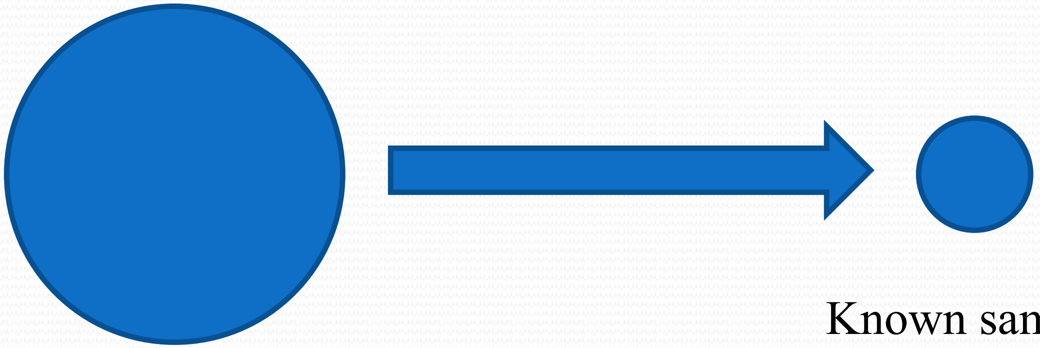
Probability



What kind of samples can occur?

Known Population

Statistics



Known sample

What kind of population
could it have come from?

Acceptance Sampling

- Suppose that an integrated circuit used in a pacemaker is supplied by a sub-contractor, with a stipulation that no more than 1% of the chips shall be defective.
  - Long-run true percent defective should be $\pi \leq 1\%$
- If 50 chips are sampled at random from a shipment of 10,000 chips, and if the defective rate is actually 1%, then what is the chance that none of the 50 chips is defective? What is the chance that exactly 1 of the 50 chips is defective?
  - If long-run true defective rate is $\pi = 1\%$
  - Sample 50 chips. Number defective out of 50 is $x = 0, 1, \ldots$
  - $P(x = 0 \mid \pi = 0.01) = ?$
  - $P(x = 1 \mid \pi = 0.01) = ?$

Acceptance Sampling

- If the shipment is rejected whenever any of the 50 chips is defective, then how often is a shipment rejected when it is in spec?
  - $P(x > 0 \mid \pi = 0.01) = ?$
  - You want this to be small
- If the shipment actually has a defective rate of 2%, how likely is it to be rejected.
  - $P(x > 0 \mid \pi = 0.02) = ?$
  - You want this to be large

Device Performance

- A pacemaker is supposed to deliver a pulse of about 90mV.
  - If average voltage across units is 90 mV and
  - Standard deviation across units is 3 mV,
  - How likely is it that a sample of 10 such pacemakers would have an average voltage above 92 mV?
- We test a sample of 10 pacemakers from a production process
  - The sample mean is 91 mV
  - The sample standard deviation is 2.5 mV
  - What is the range of likely values for the average voltage of the whole population?

Probabilistic Experiment

- The set of possible outcomes of a probabilistic experiment is called the *sample space*.
- If we roll a die, the set of possible outcomes can be represented as {1, 2, 3, 4, 5, 6}
- If we toss a coin, the set of possible outcomes can be represented as {H, T}
- If we measure the resistance of a circuit with an ohmmeter, the set of possible outcomes can be represented as $[0, \infty)$

Events

- An *event* is a subset of the sample space. The event could be said to happen if the outcome of the experiment falls into that subset.
- When rolling a die, the event that the number is even is {2, 4, 6} which is a subset of {1, 2, 3, 4, 5, 6}
- The event that a resister is in spec could be (1.20, 1.30).
- If we toss a coin three times, the sample space is {TTT, TTH, THT, THH, HTT, HTH, HHT, HHH}. The event that there are more heads than tails is {THH, HTH, HHT, HHH}

S is a sample space

A and B are subsets of S; that is, events

$A \subset S,\ B \subset S$

The *union* of A and B is the set of outcomes

that occur in A or in B or both

$A \cup B$

The *intersection* of A and B is the set of outcomes

that occur in both A and B

$A \cap B$

The *complement* of A is the set of outcomes

that are not in A

$A^C = \overline{A} = \tilde{A}$

A and B are *mutually exclusive* if the have no outcomes in common.

$A \cap B = \varnothing$

A die is rolled with sample space $\{1,2,3,4,5,6\}$

A = the number is even = $\{2,4,6\}$

B = the number is less than 5 = $\{1,2,3,4\}$

C = the number is 3 = $\{3\}$

$A \cup B = \{1,2,3,4,6\}$

$A \cup C = \{2,3,4,6\}$

$B \cup C = B = \{1,2,3,4\}$

$C \subset B$

$A \cap B = \{2,4\}$

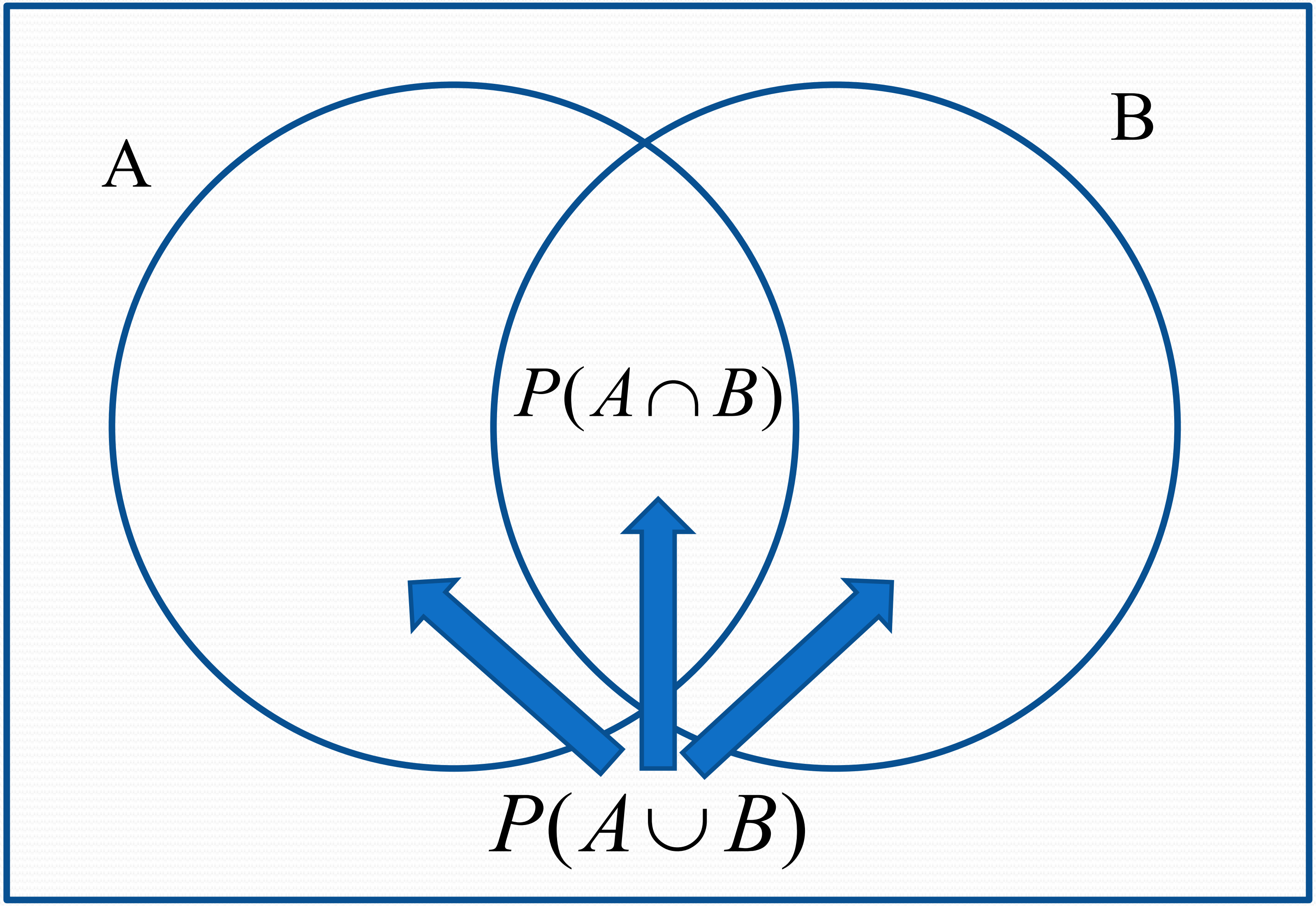
$A \cap C = \varnothing$

A and C are mutually exclusive

$B \cap C = C = \{3\}$

Probability

- In a probabilistic experiment, each event has a number assigned called its probability, which is the likelihood that the event will occur.
- If A is an event, then the probability of A occurring is denoted $P(A)$.
- If the experiment is done many times, then the fraction of times that the event occurs will be approximately $P(A)$. This is called the *law of large numbers*.


A
B
$P(A \cap B)$
$P(A \cup B)$

Laws of Probability

1. $P(S) = 1$, where S is the sample space.
2. $0 \leq P(A) \leq 1$, for any event $A \subset S$.
3. If A and B are mutually exclusive, so that $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$

Note that 3 is only true if A and B are mutually exclusive.

It is not true for all pairs of events A and B.

We will see that (always)

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

Laws of Probability

1. $P(S) = 1$, where S is the sample space.

The probability of something is a number between 0 and 1. The sample space is the set of all possible outcomes. The probability that an outcome will be in S is 1 (100%) because S contains everthing that could possibly happen, even if unlikely.

Laws of Probability

2. $0 \leq P(A) \leq 1$, for any event $A \subset S$.

The set A is a set of possible outcomes.

It could consist of all of S, in which case its probability is 1.

It could be empty ($A = \emptyset$), in which case its probability is 0.

Otherwise, its probability must be between 0 and 1.

Laws of Probability

3. If A and B are mutually exclusive, so that

   $A \cap B = \varnothing$, then

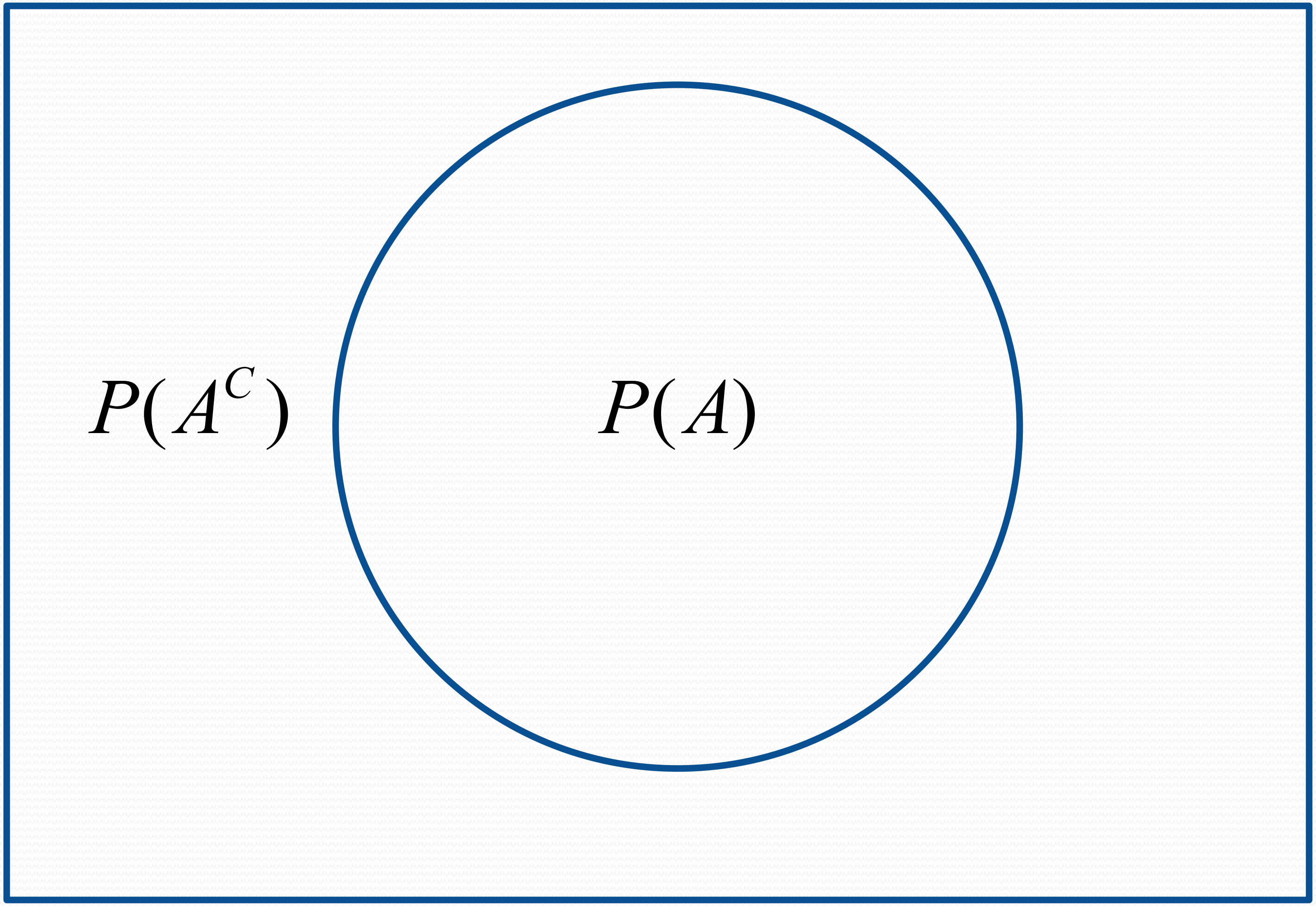
   $P(A \cup B) = P(A) + P(B)$

Note that 3 is only true if A and B are mutually exclusive.

It is not true for all pairs of events A and B.

We will see that (always)

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



$P(A^C)$
$P(A)$

If A is an event, then

$$P(A^C) = 1 - P(A)$$

If $\varnothing$ is the empty set, then

$$P(\varnothing) = 0$$

If a sample space has N equally likely outcomes,
and if an event A has k elements of the sample space, then

$$P(A) = \frac{k}{N}$$

These are all consequences of the three laws.

An extrusion die is used to produce aluminum rods. Specifications are given for the length and diameter of the rods. For each rod, the length is classified as too short, too long, or OK, and the diameters is classified as too thin, too thick, or OK. In a **population** of 1000 rods, the number of rods in each class are as follows:

	Diameter		
Length	Too Thin	OK	Too Thick
Too Short	10	3	5
OK	38	900	4
Too Long	2	25	13

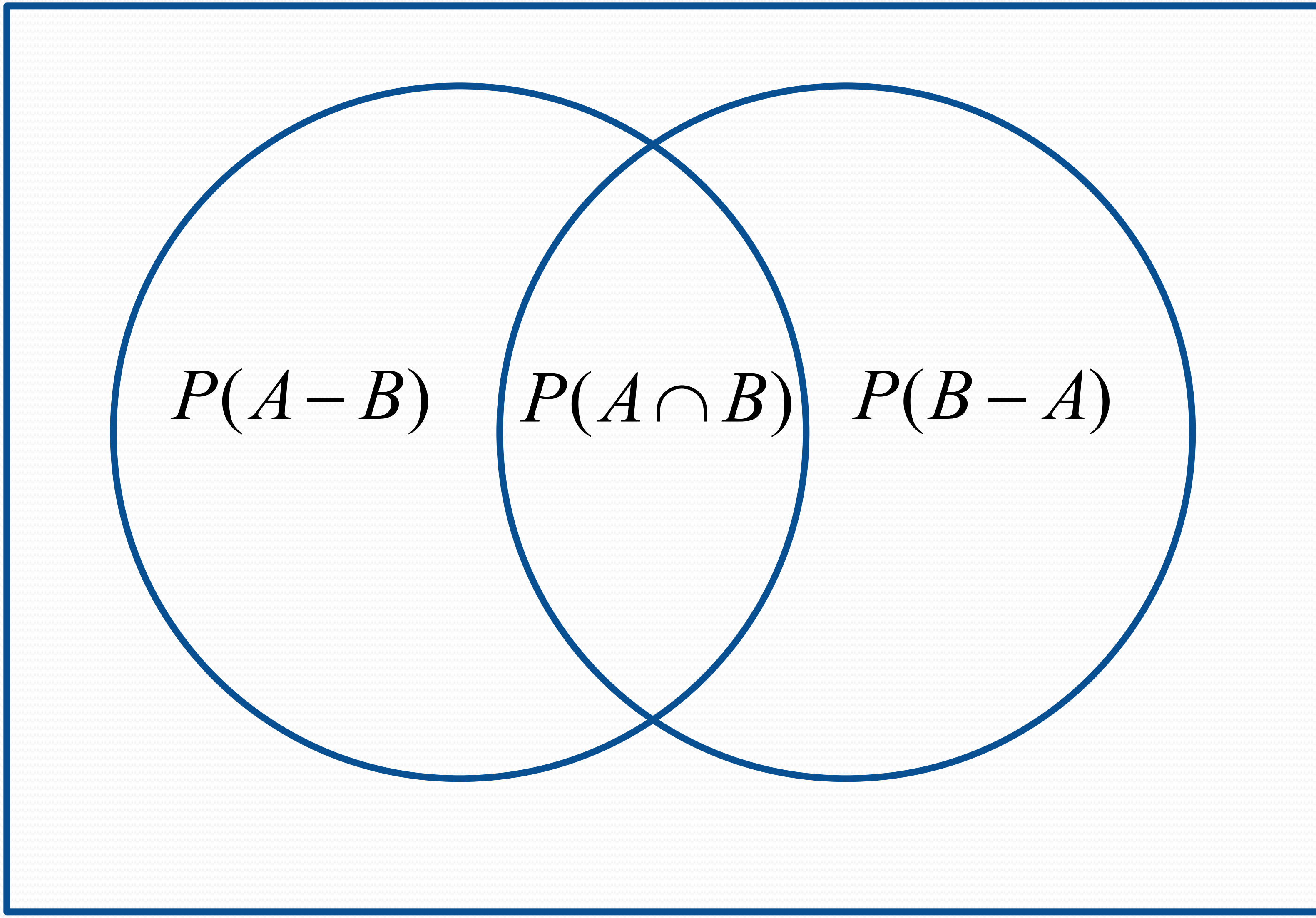
1. What is the probability that a randomly chosen rod from this population is too short?
   $(10 + 3 + 5)/1000 = 18/1000 = 0.018$

If A and B are sets (subsets of the sample space S) then

$A - B$ is the set of things in A but not in B.

$B - A$ is the set of things in B but not in A.

$A \cap B$ is the set of things in both A and B.



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

The Addition Rule

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

$A - B = A \cap B^C$ Definition

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

$A \cup B = (A - B) \cup (B - A) \cup (A \cap B)$ as in last Venn Diagram

$P(A \cup B) = P(A - B) + P(A \cap B) + P(B - A)$ because sets are disjoint

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

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

An extrusion die is used to produce aluminum rods. Specifications are given for the length and diameter of the rods. For each rod, the length is classified as too short, too long, or OK, and the diameters is classified as too thin, too thick, or OK. In a population of 1000 rods, the number of rods in each class are as follows:

	Diameter		
Length	Too Thin	OK	Too Thick
Too Short	10	3	5
OK	38	900	4
Too Long	2	25	13

If a rod is sampled at random, what is the probability that it is either too short or too thick (or both)?

(10 + 3 + 5 + 4 + 13)/1000 = 35/1000 = 0.035

18/1000 + 22/1000 − 5/1000 = 35/1000

Counting Methods

- If one operation can be done in n_1 distinct ways, and a second subsequent operation can be done in n_2 distinct ways, then the sequence of two operations can be done in $n_1 \times n_2$ distinct ways.
- If we have n distinct objects, then the number of ways we can arrange them in order is $n! = n(n-1)(n-2) \ldots (3)(2)(1)$, called "*n* factorial".
- Suppose there are 5 machining operations to be done on a part and they can be done in any order. To determine the most efficient order, we have to consider $5! = 120$ possibilities. We can choose the first operation in 5 ways, the second from any of the remaining 4, the third from any of the remaining 3, etc., so the total number of ways is $5 \times 4 \times 3 \times 2 \times 1 = 120$.

Counting Methods

- If we have n distinct objects and want to choose k of them and arrange them in some order, the number of ways to do this is $n(n-1)(n-2)(n-k+1)$.
- This is $n!$, but leaving out the terms from $(n-k)!$, so this is $n!/(n-k)!$

Counting Methods

- If we have n distinct objects and want to choose k of them, but it does not matter what the order is, then we can get this from the previous slide where we choose the k items in order, but there are many orders that get the same set of k objects; in fact there are $k!$ such orders.
- So the set of choices of k things out of n, or the combinations of k things out of n, or "n choose k" is $n!/[k!(n - k)!]$

The number of permutations of n things is

$$P(n) = n!$$

The number of permutations of k things chosen out of n is

$$P(n,k) = \frac{n!}{(n-k)!}$$

The number of combinations of k things out of n, or n-choose-k is

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

This is also called the binomial coefficient because it occurs in algebra when we raise a binomial (two-term expression) to a power

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

Conditional Probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

This only makes sense if

$P(B) \neq 0$

Conditional Probability

- 900/1000 rods are ok on length and ok on diameter
- A = ok on length, B = ok on diameter
- P(A) = 942/1000 = 0.942
- P(B) = 928/1000 = 0.928
- P(A and B) = 900/1000 = 0.900
- P(A|B) = probability of A given B?
- 928 rods satisfy condition B on diameter; Of those, 900 also satisfy condition A
- P(A|B) = 900/928 = 0.970
- P(B|A) = 900/942 = 0.955

	Diameter		
Length	Too Thin	OK	Too Thick
Too Short	10	3	5
OK	38	900	4
Too Long	2	25	13

Throw two dice

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Sum	2	3	4	5	6	7	8	9	10	11	12
Number	1	2	3	4	5	6	5	4	3	2	1
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Sum	2	3	4	5	6	7	8	9	10	11	12
Number	1	2	3	4	5	6	5	4	3	2	1
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

- P(Sum = 6) = 5/36 = 0.139
- P(two dice have the same number) = 6/36 = 0.167
- P(same number | sum = 6) = (1/36)/(5/36) = 1/5 = 0.20
- P(same number | sum = 4) = (1/36)/(3/36) = 1/3 = 0.33
- P(same number | sum = 2) = (1/36)/(1/36) = 1/1 = 1.00

Conditional Probability

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

(Statistical) Independence

A and B are *statistically independent* iff

$$P(A \mid B) = P(A)$$

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

A and B are statistically independent iff

the occurance of B or not provides no information

about whether A will occur

Mutually Exclusive vs. Statistically Independent

Many students confuse these two very different concepts

A and B are statistically independent iff

$P(A \mid B) = P(A)$

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

A and B are mutually exclusive iff

$A \cap B = \varnothing$ so that

$\mathrm{P}(\mathrm{A} \cap B) = 0$

So these are very different concepts.

A and B are **never both** statistically independent and mutually exclusive

(unless $P(A) = 0$ or $P(B) = 0,$ which is uninteresting)

A and B are statistically independent iff

$P(A \mid B) = P(A)$

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

The occurance of A provides **no information** about $P(B)$

A and B are mutually exclusive iff

$A \cap B = \varnothing$ so that

$\mathrm{P(A} \cap B) = 0$

If A occurs, then B cannot occur, so the occurrance of A provides **complete information** about $P(B)$

The Multiplication Rule

$$P(A \cap B) = P(B)P(A \mid B) \qquad P(B) \neq 0$$
$$P(A \cap B) = P(A)P(B \mid A) \qquad P(A) \neq 0$$

The Multiplication Rule for Statistically Independent Events

If A and B are statistically independent events then

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

	Diameter		
Length	Too Thin	OK	Too Thick
Too Short	10	3	5
OK	38	900	4
Too Long	2	25	13

- Pr(Too Long) = 40/1000 = 0.040
- Pr(Diameter OK) = 928/1000 = 0.928
- Pr(Too Long|Diameter OK) = 25/928 = 0.0269 ≠ 0.04
  - Not independent
- Pr(Too Thin) = 50/1000 = 0.050
- Pr(Too Thin|Too Long) = 2/40 = 0.05 = Pr(Too Thin)
- Pr(Too Long|Too Thin) = 2/50 = 0.04 = Pr(Too Long)
  - Independent

Statistical Independence in Practice

- We can calculate when certain events are statistically independent
- If we pick one card from a 52-card deck, the events {Ace} and {Spade} are statistically independent (4/52 = 1/13)
- Mostly in practice, the statistical independence is a consequence of the physical setup
- If we grow up three plates from frozen cells and measure the number of cells after 18 hours, we can consider the numbers in the three plates to be statistically independent.
- We can then use the multiplication rule to find the probabilities of various outcomes.

The Law of Total Probability

Suppose that a set of events partitions the sample space:

$A_1, A_2, \ldots, A_n$

Mutually Exclusive

$A_i \cap A_j = \varnothing$ for all i, j

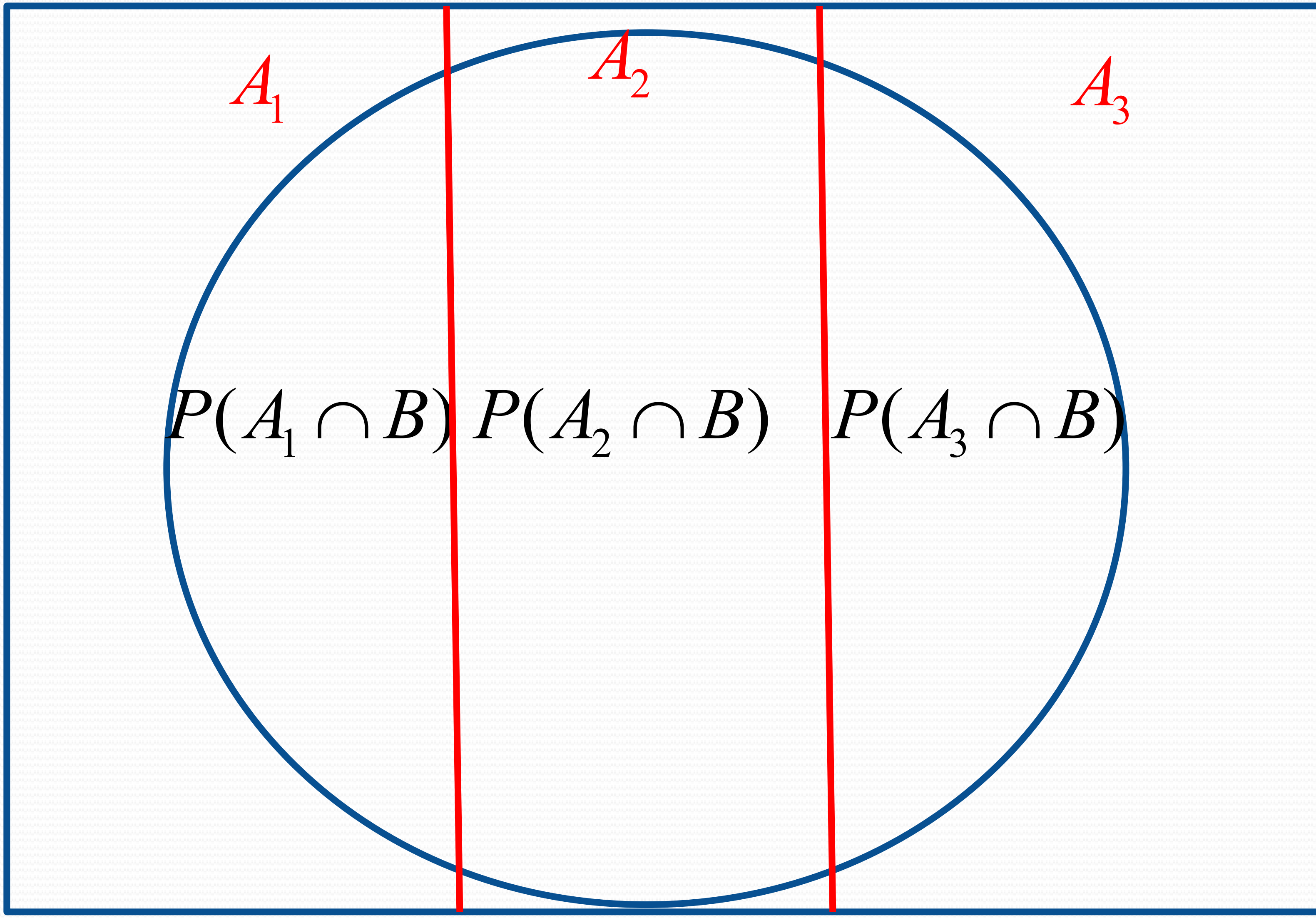
Collectively Exhaustive

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

If B is any event, then

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots P(A_n \cap B)$$

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \cdots + P(B \mid A_n)P(A_n)$$

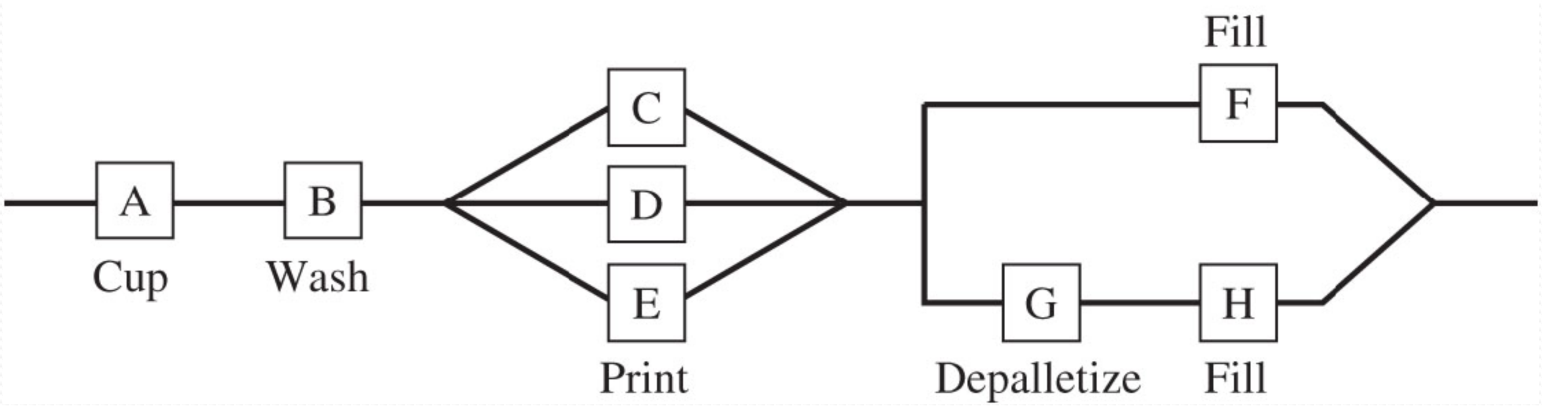

A_1
A_2
A_3
$P(A_1 \cap B)$
$P(A_2 \cap B)$
$P(A_3 \cap B)$

Reliability

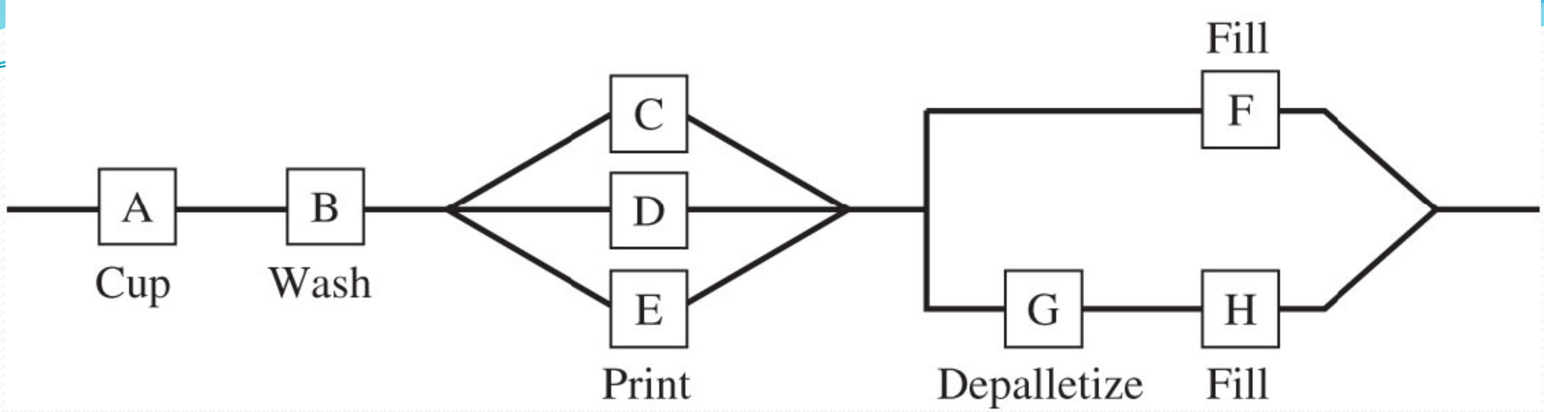
- A lab has three printers. Each printer has a chance of 98% of functioning on a given day and the failures are statistically independent. What is the chance that at least one printer is operating?
- The complement of the event that at least one is working is that none of them is working (number working is 0, 1, 2, or 3).
- The chance that printer 1 is not working is 0.02 and the same for printers 2 and 3.
- Since they are statistically independent, the chance that all three are not working is the product of the probabilities (0.02)(0.02)(0.02) = .000008 or 1 chance in 125,000.
- The probability that at least one is working is 0.999992.

- A production process for a biomedical device requires that two devices A and B both be working. If the probability that device A works is 0.98 and the probability that device B works is 0.94, and if failures are statistically independent, what is the chance that the system is working?
- P(A and B are working) = P(A is working)P(B is working) = (0.98)(0.94) = 0.9212.
- If we only needed one of the two to be working, what is that probability?
- Chance that both are failed is (0.02)(0.06) = 0.0012, so the chance that at least one is working is 1 – 0.0012 = 0.9988.

Manufacture of Aluminum Cans

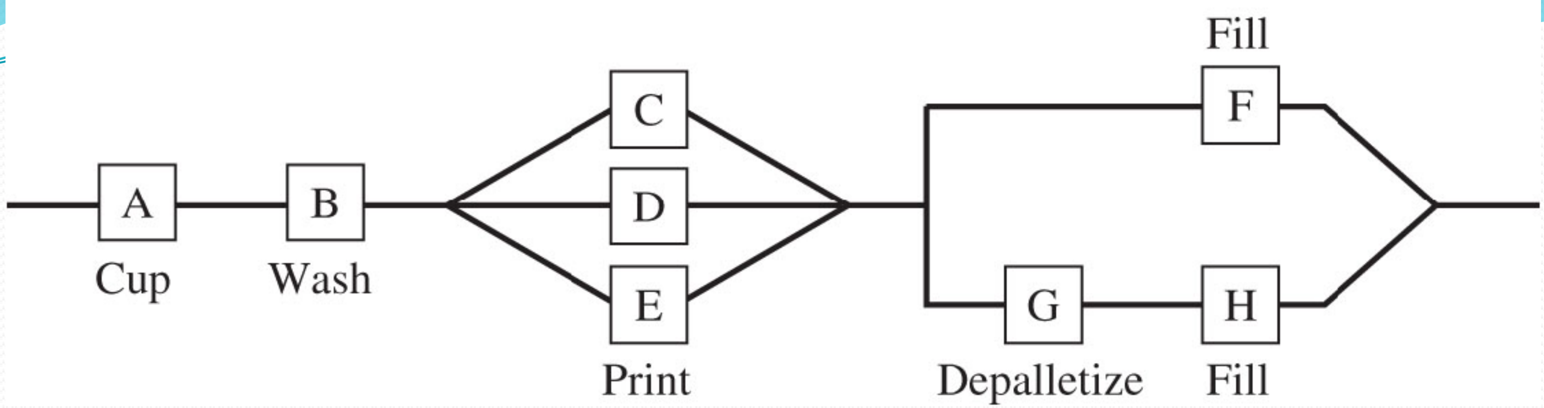


- Aluminum sheets 0.25mm thick are "cupped' into cans open at the top, then washed.
- One of three printers makes the label. They are placed in pallets of 7,140 cans.
- Two types of fill lines. In both cases, the cans are filled and sealed



- For each component, we know the probability that it will function on any given day, and we assume that the failures are statistically independent.
- The probability of functioning is given in the table below.

A	B	C	D	E	F	G	H
0.995	0.99	0.95	0.95	0.95	0.90	0.90	0.98



A	B	C	D	E	F	G	H
0.995	0.99	0.95	0.95	0.95	0.90	0.90	0.98

- P(Cup + Wash) = (0.995)(0.99) = 0.98505
- P(Print) = 1 – (0.05)(0.05)(0.05) = 0.999875
- P(Depalletize + Fill) = (0.90)(0.98) = 0.882
- P(Fill System) = 1 – (0.118)(0.10) = 0.9882
- P(System) = (0.98505)(0.999875)(0.9882) = 0.9733
- P(System Fails) = 1 – 0.9733 = 0.0267 or 1/37.45
- System fails a little less often than once a month.

Summary

1. $P(S)=1$, where S is the sample space.
2. $0 \le P(A) \le 1$, for any event $A \subset S$.
3. If A and B are mutually exclusive, so that $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$

$$P(A^C) = 1 - P(A)$$

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

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$ This only makes sense if $P(B) \neq 0$

A and B are statistically independent iff

$$P(A \mid B) = P(A)$$

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

If $A_1, A_2, \ldots, A_n$ are mutually exclusive and collectively exhaustive and B is any event then

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots P(A_n \cap B)$$

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \cdots + P(B \mid A_n)P(A_n)$$