Probability

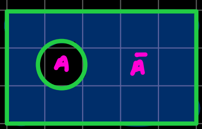
- **Probability:** Numerical measure of likelihood that an event will occur
- **sample space:** Set of all experimental outcomes
  - → An experimental outcome is called a sample point
- **multi-step experiments:** Refer to SI notes

- **Counting Techniques:**

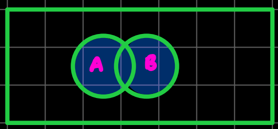
① $n!$
② $(n-1)!$
③ ${}^{n}C_{r} = \dfrac{n!}{(n-r)!}$
④ ${}^{n}P_{r} = \dfrac{n!}{r!(n-r)!}$

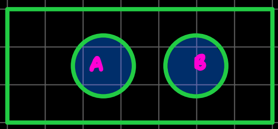
- **Events:** An event is a collection of sample points
  - → $P(E)$ = Sum of Prob of all sample points

- **Complement:** $P(\bar{A}) = 1 - P(\bar{A})$



- **Addition Law:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ – non-mutually exclusive
  - → $P(A \cup B) = P(A) + P(B)$ – mutually exclusive





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

- **Multiplication Law:** $P(A \cap B) = P(B)\,P(A \mid B)$
  - $P(A \cap B) = P(A)\,P(B \mid A)$
  - → $P(A \cap B) = P(A)\,P(B)$ – For checking independant events
  - * $P(A \cap B \cap C) = P(A)\,P(B \mid A)\,P(C \mid A \cap B)$
  - $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$

TABLE 4.5 JOINT PROBABILITY TABLE FOR PROMOTIONS

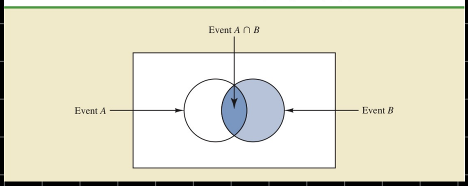
Joint probabilities appear in the body of the table.

	Men (M)	Women (W)	Total
Promoted (A)	.24	.03	.27
Not Promoted (A^c)	.56	.17	.73
Total	.80	.20	1.00

Marginal probabilities appear in the margins of the table.

- **Conditional Prob:** $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

FIGURE 4.8 CONDITIONAL PROBABILITY $P(A \mid B) = P(A \cap B)/P(B)$



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

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

- If $P(A \mid B) = P(A)$, events are independant, otherwise dependent

- **Mutually Exclusive:** Events that can not happen simultaneously
  - → $P(A \cap B) = 0$

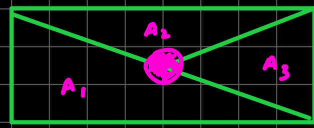
- **Exhaustive Events:** Union of mutually exclusive events from the sample space

Baye's Theorem

① $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$

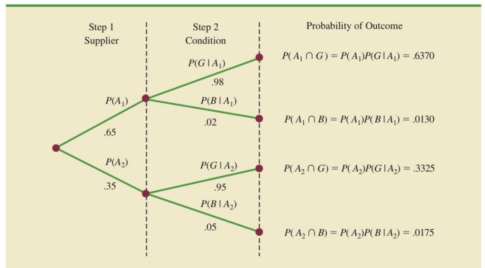
$= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)$

$\rightarrow \sum_{i=1}^{n} P(A_i)P(B|A_i)$



② $P(A_i | B) = \dfrac{P(A_i \cap B)}{P(B)} = \dfrac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$

FIGURE 4.11 PROBABILITY TREE FOR TWO-SUPPLIER EXAMPLE



Tabular Approach

A tabular approach is helpful in conducting the Bayes' theorem calculations. Such an approach is shown in Table 4.7 for the parts supplier problem. The computations shown there are done in the following steps.

Step 1. Prepare the following three columns:
Column 1—The mutually exclusive events A_i for which posterior probabilities are desired
Column 2—The prior probabilities $P(A_i)$ for the events
Column 3—The conditional probabilities $P(B | A_i)$ of the new information B given each event

Step 2. In column 4, compute the joint probabilities $P(A_i \cap B)$ for each event and the new information B by using the multiplication law. These joint probabilities are found by multiplying the prior probabilities in column 2 by the corresponding conditional probabilities in column 3; that is, $P(A_i \cap B) = P(A_i)P(B | A_i)$.

Step 3. Sum the joint probabilities in column 4. The sum is the probability of the new information, $P(B)$. Thus we see in Table 4.7 that there is a .0130 probability that the part came from supplier 1 and is bad and a .0175 probability that the part came from supplier 2 and is bad. Because these are the only two ways in which a bad part can be obtained, the sum .0130 + .0175 shows an overall probability of .0305 of finding a bad part from the combined shipments of the two suppliers.

Step 4. In column 5, compute the posterior probabilities using the basic relationship of conditional probability.

$$P(A_i | B) = \frac{P(A_i \cap B)}{P(B)}$$

Note that the joint probabilities $P(A_i \cap B)$ are in column 4 and the probability $P(B)$ is the sum of column 4.

TABLE 4.7 TABULAR APPROACH TO BAYES' THEOREM CALCULATIONS FOR THE TWO-SUPPLIER PROBLEM

(1) Events A_i	(2) Prior Probabilities $P(A_i)$	(3) Conditional Probabilities $P(B \mid A_i)$	(4) Joint Probabilities $P(A_i \cap B)$	(5) Posterior Probabilities $P(A_i \mid B)$
A_1	.65	.02	.0130	.0130/.0305 = .4262
A_2	.35	.05	.0175	.0175/.0305 = .5738
	1.00		$P(B)$ = .0305	1.0000

Q- Select 3 from 15 employees. Of 15, 8 are cyberspace experts and 7 are general IT members. If selected at random

a) Prob that all selected are cyberspace experts
b) At least 1 IT staff member is selected

a) Total = ${}^{15}C_3$ Prob = $\dfrac{{}^{8}C_3}{{}^{15}C_3} = \dfrac{8}{65}$

Cyberspace = ${}^{8}C_3$

b) IT _ _ ${}^{7}C_1 \times {}^{8}C_2 = 196$
IT IT _ ${}^{7}C_2 \times {}^{8}C_1 = 168$
IT IT IT ${}^{7}C_3 = 35$
$\overline{399}$

Prob = $\dfrac{399}{{}^{15}C_3} = \dfrac{57}{65}$

Q- 0.25 — defect in braking system
0.18 — defect in transmission
0.17 — defect in fuel system
0.4 — defect in other area   Find Prob

a) Brakes or fueling system, if prob of defects in both systems is 0.15.
b) no defects in either brakes or fueling system

a) $P(\text{Brakes} \cup \text{Fuel}) = P(B) + P(F) - P(B \cap F)$
$= 0.25 + 0.17 - 0.15 = 0.27$

b) $1 - 0.27 = 0.73$

Q- Two OS : X, Y

- 40% of all bugs in X
- 30% of all bugs in Y
- 15% of all bugs in both

a) Prob that a bug report is related to at least 1 of 2 versions

a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.4 + 0.3 - 0.15 = \frac{11}{20}$

Q- - 72% for grocery
- 56% for food court
- 60% clothing
- 50% grocery and food court
- 45% grocery and clothing
- 30% visit all 3

a) Prob for clothing given that grocery
b) Prob for food court, given that clothing and grocery

a) $P(C \mid G) = \frac{P(C \cap G)}{P(G)} = \frac{0.45}{0.72} = \frac{5}{8}$

b) $P(F \mid C \cap G) = \frac{P(F \cap C \cap G)}{P(C \cap G)} = \frac{0.3}{0.45} = \frac{2}{3}$

Q-

	Quality	Cost / Convenience	other	
Full-Time	421	393	76	890
Part-Time	400	593	46	1039
	821	986	122	1929

a) Develop joint probability table
b) cal marginal prob of school quality, cost & convenience
c) If a student goes full time what is prob that school quality is the reason
d) If a student goes part-time, what prob is that cost & convenience is the reason
e) Let A= Full-Time, and B= Reason is school quality. Are A & B independant

a,b)

	Quality	Cost & Convinience	other	
Full-Time	421/1929	393/1929	76/1929	0.46137
Part-Time	400/1929	593/1929	46/1929	0.53862
	0.4256	0.51111	0.06324	1

c) $P(Q \mid F) = \frac{P(Q \cap F)}{P(F)} = \frac{\frac{421}{1929}}{\frac{421+400}{1929}} = \frac{421}{821}$

d) $P(C \mid P) = \frac{\frac{593}{1929}}{\frac{593+393}{1929}}$

$= \frac{593}{986}$

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

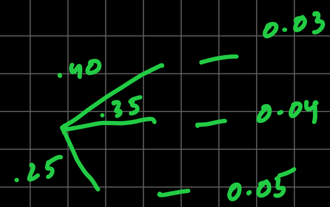
$\frac{421}{1929} = \frac{421+393+76}{1929} \times \frac{421+400}{1929}$

$\frac{421}{1929} = \frac{890}{1929} \times \frac{821}{1929}$

$0.22 \neq 0.20$, hence not independant

Q- A, B, C – 3 companies

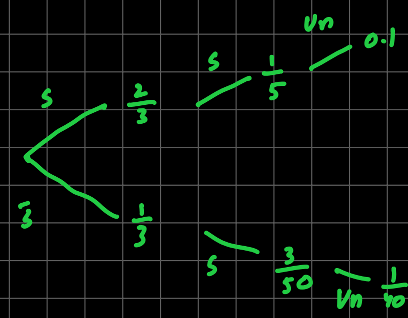
A- provide 40 % of laptops , defect rate 3%
B- 35 % laptops, defect rate 4 %
C- 25 % laptops, defect rate 5 %



a) Prob of getting a defective laptop
b) Prob of being from company C , given that laptop is defective

a) $P(D) = 0.4 \times 0.03 + 0.35 \times 0.04 + 0.25 \times 0.05 = 0.0385$

b) $P(D \mid C) = \dfrac{0.25 \times 0.05}{0.0385} = \dfrac{25}{77}$



Q- - set of 2000 spam messages
- Set of 1000 non spam messages
- word 'stock' appears in 400 spam messages , and 60 msgs not spam
- word 'undervalued' appears in 200 spam msgs , and in 25 non-spam msgs

a) Estimate prob , a message containing both stock and undervalued is spam.
b) will the message be rejected as spam when threshold is set at 0.9

a) $P(\text{spam} \mid \text{stock} \cap \text{under}) = \dfrac{P(\text{spam} \cap \text{stock} \cap \text{under})}{P(\text{stock} \cap \text{under})}$

$P(\text{Stock} \cap \text{Under}) = \frac{2}{3} \times \frac{1}{5} \times 0.1 + \frac{1}{3} \times \frac{3}{50} \times \frac{1}{40} = \frac{83}{6000}$

$P(\text{Spam} \mid \text{stock} \cap \text{Under}) = \dfrac{\frac{2}{3} \times \frac{1}{5} \times 0.1}{83 \div 6000} = 0.96385$

(b) Rejected as spam