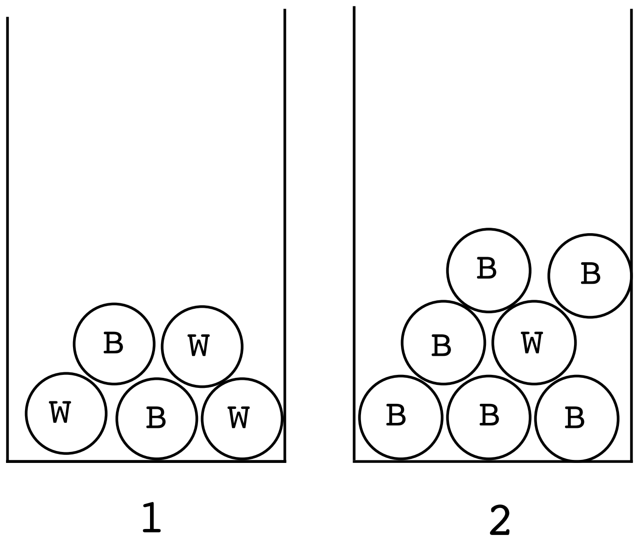
Lecture 3
More manipulation rules.

Consider the following set-up

Throw a coin. If H then pick ball from bin 1 (3W,2B).
If T then pick from bin 2 (1W,6B)



$P(H) = ½$ $\quad$ $P(T) = 1/2$

$P(W \mid H) = 3/5$ $\quad$ $P(W \mid T) = 1/7$

The Partition Law

What is P(W)? (P(W ball chosen))

We know:

$P(H) = 1/2$	$P(T) = 1/2$
$P(W \mid H) = 3/5$	$P(W \mid T) = 1/7$

W happens if and only if (H then W) or (T then W) thus

$$
\begin{aligned}
P(W) &= P((H\&W) \text{ or } (\mathrm{T\&W})) \\
&= \mathrm{P(H\&W) + P(T\&W)} \\
&= \mathrm{P(H)P(W|H) + P(T)P(W|T)} \\
&= \frac{1}{2}\times\frac{3}{5} \quad + \quad \frac{1}{2}\times\frac{1}{7} = \frac{13}{35}
\end{aligned}
$$

This is an example of the *Partition Law.*

The Partition Law:

Let A be any event.

Let $x_1, x_2, \ldots, x_n$ be all the possible outcomes of another experiment.

Then

$$P(A) = \sum_{i=1}^{n} P(x_i)\ P(A \mid x_i)$$

Example

Three robotic stations A,B, and C assemble identical mother boards (MBs)

A batch of 200 MBs was produced

100 by station A,
60 by station B
40 by station C.

Thus : $P(A)=0.5$, $P(B)=0.3$, $P(C)=0.2$.

The probabilities of a fault on a MB depends on the machine that produced it.

$P(faulty|A) =0.04$,
$P(faulty|B)=0.02$,
$P(faulty|C)=0.01$.

P(random MB from batch faulty) =

$$0.04*0.5 + 0.02*0.3+0.01*0.2 =$$

$$0.020 + 0.006 + 0.002 = 0.028.$$

Continued:

A simple test has a 99% chance of detecting a faulty MB. However the test will report an OK MB as faulty with probability 0.05.

P(reported as faulty) =
P(rep | faulty) *P(faulty) + P(rep | OK) * P(OK)=
0.99 * 0.028 + 0.05 * 0.972 = 0.07632.

Note that only 2.8% are actually faulty, over 7.6% are reported as faulty by the test!

Address this in the next section.

Bayes' Theorem

Want P(A | B). Apply definition of conditional probability twice

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

$$= \frac{P(A)P(B|\,A)}{P(B)}$$

This equality is the basis of **........**

Bayes' Law:

Let A be any event.

Let $x_1, x_2, \ldots, x_n$ be all the possible outcomes of another event. Then

$$P(x_j \mid A) = \frac{P(x_j)\, P(A|\, x_j)}{\sum_{i=1}^{n} P(x_i)\, P(A|\, x_i)}$$

Proof:

$$P(x_j \mid A) = \frac{P(x_j \text{ and } A)}{P(A)}$$

$$= \frac{P(x_j)P(A \mid x_j)}{P(A)}$$

$$= \frac{P(x_j)P(A \mid x_j)}{\sum_{i=1}^{n} P(x_i)P(A \mid x_i)}$$

In the MB example we can find out which station to blame for a defective:

$$P(A \mid faulty) = \frac{0.04 * 0.5}{0.028} = \frac{0.02}{0.028} \approx 0.714$$

Note P(A)=0.5 , finding the evidence that the MB was faulty has increased the probability that it comes from A to 0.71, because the A station has the worst record for faulty MBs.

Exs. Compute P(B|faulty) and P(C|faulty) .

The test

With the test we would like to know:

$$P(OK \mid repOK) = \frac{P(repOK \mid OK) * P(OK)}{P(repOK)} =$$

$$= \frac{0.95 * 0.972}{0.92468} = 0.9986$$

So the test works quite well only 0.14% of the boards passed are faulty. – This is called the *sensitivity* of the test.

0.99 * 0.028 + 0.05 * 0.972 = 0.07632.

$$p(faulty \mid repfaulty) = \frac{0.99 * 0.028}{0.07632} = 0.3632$$

So only 36% of MBs reported faulty actually are, thus 64% are OK!. This is a common problem with screening tests – too many false positives . This probability is called the test specificity.

Bayes' thm is used for modifying the probability of an outcome given some evidence. An extension of the technique is in fact used in the current SPAM filter software.

Data:

P(SPAM| property A) =