

## Lecture 2

Last time:

Given a set of events with are **mutually exclusive** and **equally likely**,

$$P(E) = \frac{n(E)}{N}.$$

*Example: card games*

$$\text{Number of different 5-card poker hands} = \binom{52}{5} = 2,598,960$$

$$\text{Number of different 13-card bridge hands} = \binom{52}{13} = 635,013,559,600$$

Poker probabilities:

$$1. \quad P(\text{one pair}) = \frac{13 \binom{4}{2} 12 \binom{4}{1} 11 \binom{4}{1} 10 \binom{4}{1}}{\binom{52}{5} 3!} = 0.423$$

$$2. \quad P(\text{two pair}) = \frac{13 \binom{4}{2} 12 \binom{4}{2} 11 \binom{4}{1}}{\binom{52}{5} 2!} = 0.0476$$

## Independence

$P(AB) = P(A)P(B)$  when A,B independent

$P(AB) = P(A)P(B|A)$  when A,B dependent

Definition of independence of events:  $P(A_1 A_2 A_3 \dots A_m) = P(A_1)P(A_2) \dots P(A_m)$  for all m (pairwise, threewise, etc.)

Definition of conditional probability:  $P(B|A) = \frac{P(AB)}{P(A)}$

$P(B|A)$  restricts our attention to the situation where A has already occurred.

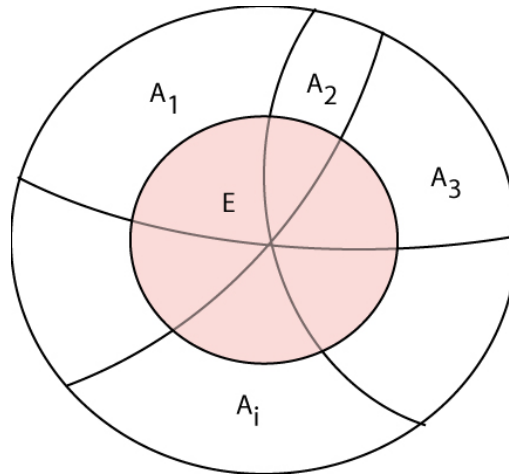
If A,B are independent

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

Two useful results:

1.  $P(ABCD\dots) = P(A)P(B|A)P(C|AB)P(D|ABC)\dots$  Derive this by letting  $A=CD$ . Then  $P(BCD) = P(CD)P(B|CD) = P(C)P(D|C)P(D|CD)$
2. If  $A_1, A_2, \dots$  is a set of **mutually exclusive** and **collectively exhaustive** events, then

$$P(E) = P(EA_1) + P(EA_2) + \dots + P(EA_n) = P(A_1)P(E|A_1) + P(A_2)P(E|A_2) + \dots + P(A_n)P(E|A_n)$$



$P(A+B) = P(A) + P(B) - P(AB)$  must subtract off  $P(AB)$  because it is counted twice by the first two terms of the RHS.

$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$  (for three events).

If four events,  $P(A+B+C+D)$  would be the sum of probabilities of the single events, minus the probability of the joint events taken two at a time, plus the probability of the joint events taken three at a time, minus the probability of the joint event taken four at a time.

**Conclusion:** If events are mutually exclusive, express result in terms of the sum of events; if independent, in terms of joint events. Note that this reduces to just the sum of the probabilities of the events if the events are mutually exclusive.

### *Example: Simplex system*

Probability of component failure =  $P(F_i)$

$P(\text{system failure}) = P(\text{any component fails}) = P(F_1 + F_2 + \dots + F_n)$

Assume component failures are **independent**. Take advantage of this by working with the complementary event.

$$P(\text{system failure}) = 1 - P(\text{system works}) = 1 - P(W_1 W_2 \dots W_n) = 1 - \prod_i P(W_i) = 1 - \prod_i (1 - P(F_i))$$

where  $P(W_i) = 1 - P(F_i)$ .

**Example: *n*-bit message**

$p_e$  - bit error probability

A system may be able to tolerate up to  $k$  **independent** errors and still decode the message. Compute the probability of  $k$  errors.

$\square \square \dots \square \square$

$$P(k \text{ errors}) = \binom{n}{k} p_e^k (1 - p_e)^{n-k}$$

**Conditional Independence**

Sometimes events depend on more than one random quantity, some of which may be common to two or more events, the others different and independent. In such cases the concept of conditional independence is useful.

**Example: Two measurements of the same quantity,  $x$ , each corrupted by additive independent noise.**

$$m_1 = x + n_1$$

$$m_2 = x + n_2$$

$$E_1(m_1), E_2(m_2)$$

Since  $m_1$  and  $m_2$  both depend on the same quantity,  $x$ , any events which depend on  $m_1$  and  $m_2$ ,  $E_1(m_1)$  and  $E_2(m_2)$ , are clearly not independent.

$$P(E_1 E_2) \neq P(E_1) P(E_2) \text{ in general}$$

$n_i$  are independent

$$P(AB | E) = P(A | E) P(B | E)$$

$$P(B | A) \Rightarrow P(B) \text{ if A, B are independent}$$

$$P(B | EA) = \frac{P(ABE)}{P(EA)} = \frac{P(E)P(AB | E)}{P(E)P(A | E)} = \frac{P(A | E)P(B | E)}{P(A | E)} = P(B | E)$$

But if the value of  $x$  is given, the statistical properties of  $m_1$  and  $m_2$  depend only on  $n_1$  and  $n_2$ , which are independent. Thus the events  $E_1$  and  $E_2$ , conditioned on a given value of  $x$ , would be independent. Intuitively we would then say

$$P(E_1 E_2 | x) = P(E_1 | x) P(E_2 | x)$$

In general, the statement

$$P(AB | E) = P(A | E) P(B | E)$$

is the definition of the conditional independence of the events  $A$  and  $B$ , conditioned on the event  $E$ .

If  $A$  and  $B$  are conditionally independent, given  $E$ , and  $B$  is conditioned on  $E$ , further conditioning on  $A$  should not change the probability of the occurrence of  $B$  since if  $E$  is given,  $A$  and  $B$  are conditionally independent.

## **Bayes' Theorem**

An interesting application of conditional probability. Let  $A_i$  be mutually exclusive and collectively exhaustive.

$$P(A_k | E) = \frac{P(A_k E)}{P(E)} = \frac{P(A_k E)}{\sum_i P(A_i E)} = \frac{P(A_k) P(E | A_k)}{\sum_i P(A_i) P(E | A_i)} \quad \text{Bayes' rule}$$

This relation results directly from the definition of conditional probability and cannot be questioned. But over the years, the application of this relation to statistical inference has been subjected to a great deal of mathematical and philosophical criticism. Of late the notion has been more generally accepted, and serves as the basis for a well-defined theory of estimation and decision-making.

The point of the controversy is the use of a probability function to express one's state of imperfect knowledge about something which is itself not probabilistic at all.

### *Example: Acceptance testing*

Components are manufactured by a process which is uncertain – some come out good and some bad. A test is devised such that a good component has

80% chance of passing and a bad component has 30% chance of passing. On the average, we think that the process produces 60% good components, 40% bad ones. If a component passes the test, what is the probability that it is a good one?

Note that whether a certain component is good or bad is really not probabilistic - it is either good or bad. But we don't know what it is. Our knowledge based on all prior experience, intuition, or whatever is expressed in the prior probabilities.

$$P(\text{"good"}) = P(G) = A_1 = 0.6$$

$$P(\text{"bad"}) = P(B) = A_2 = 0.4$$

$$P(P|G) = 0.8$$

$$P(P|B) = 0.3$$

$$P(G|P) = \frac{P(G)P(P|G)}{P(G)P(P|G) + P(B)P(P|B)} = \frac{0.6(0.8)}{0.6(0.8) + 0.4(0.3)} = \frac{0.48}{0.60} = 0.8$$

Suppose we have  $n - 1$  observations

$$P(A_k | E_1 \dots E_{n-1})$$

Take one more observation,  $E_n$

$$\begin{aligned} P(A_k | E_1 \dots E_n) &= \frac{P(A_k E_1 \dots E_n)}{P(E_1 \dots E_n)} \\ &= \frac{P(A_k E_1 \dots E_n)}{\sum_i P(A_i E_1 \dots E_n)} \\ &= \frac{P(E_1 \dots E_{n-1}) P(A_k | E_1 \dots E_n) P(E_n | A_k E_1 \dots E_{n-1})}{\sum_i P(E_1 \dots E_n) P(A_i | E_1 \dots E_{n-1}) P(E_n | A_i E_1 \dots E_{n-1})} \\ &= \frac{P(A_k | E_1 \dots E_{n-1}) P(E_n | A_k)}{\sum_i P(A_i | E_1 \dots E_{n-1}) P(E_n | A_i)} \end{aligned}$$

This is of the same form as the relation first written down, with  $P(A_k | E_1 \dots E_{n-1})$  in place of  $P(A_k)$ . This says that Bayes' rule can be applied repetitively to account for any number of observations if the observations are conditionally independent - conditioned on the alternatives.