

## Notes for Chapter 2 of DeGroot and Schervish

### Conditional Probabilities

Sometimes if we are given additional information we can reduce our sample space.

If we know an event  $F$  occurs, then the probability of another event  $E$  occurring may change.

Ex. It snows tomorrow, given it is snowing today.

If event  $F$  occurs what is the probability that event  $E$  occurs?

This probability is called the conditional probability of  $E$  given  $F$ , written  $P(E|F)$ .  
 $E|F$  symbolizes the event  $E$  given the event  $F$ .

Ex. Flip two coins.

$$S = \{HH, HT, TH, TT\}$$

$$P(HH) = \frac{1}{4}$$

$F$  = at least one  $H$

If we know  $F$  occurs, then our new sample space is  $S' = \{HH, HT, TH\}$

$$P(HH|F) = \frac{1}{3}$$

*The probability of getting two  $H$  is greater if we already know  $F$ .*

$G$  = the first flip is  $H$

$$S'' = \{HH, HT\}$$

$$P(HH|G) = \frac{1}{2}$$

*And greater yet if we already know  $G$ .*

Definition: If  $P(F) > 0$ , then  $P(E|F) = P(EF)/P(F)$

(Draw Venn diagram.)

What if  $A$  and  $B$  are disjoint?

If  $A$  and  $B$  are disjoint then  $P(AB) = 0$ . Hence,  $P(A|B) = 0$ . If  $B$  occurs, then  $A$  cannot occur.

Ex. What is the probability of rolling an even number with a single die, given the die roll is 3 or less?

$$S = \{1,2,3,4,5,6\}$$

$$A = \text{roll an even number} = \{2,4,6\}$$

$$B = \text{roll a 3 or less} = \{1,2,3\}$$

$$AB = \{2\}$$

$$P(A|B) = P(AB)/P(B) = P(\{2\})/P(\{1,2,3\}) = 1/3$$

What is the probability the die roll is 3 or less, given the die roll is an even number?

$$P(B|A) = P(AB)/P(A) = P(\{2\})/P(\{2,4,6\}) = 1/3$$

Ex. A family has two children. What is the probability that both are boys, given that at least one is a boy?

Let (b, g) denote the event that the older child is a boy and the younger child a girl.

$$S = \{(b, b), (b, g), (g, b), (g, g)\}.$$

Assume that all outcomes are equally likely.

E = both children are boys

F = at least one of them is a boy

$$P(E|F) = P(EF)/P(F) = P(\{(b,b)\}) / P(\{(b,b), (b,g), (g,b)\}) = (1/4) / (3/4) = 1/3$$

What is the probability that both are boys given that the younger is a boy?

E = both children are boys

G = the younger is a boy.

$$P(E|G) = P(EG)/P(G) = P(\{b,b\})/P(\{g,b\}, \{b,b\}) = (1/4) / (1/2) = 1/2$$

Ex. At a fast food restaurant, 90% of the customers order a hamburger. If 72% of the customers order a hamburger and french fries, what is the probability that a customer who orders a hamburger will also order french fries?

A = customer orders a hamburger

B = customer orders french fries

$$P(B|A) = P(AB)/P(A) = 0.72/0.9 = 0.8$$

Conditional probabilities can be used to calculate the probability of the intersection of two events.

We can rewrite  $P(E|F) = P(EF)/P(F)$ , as  $P(EF) = P(E|F)P(F)$

Note that  $P(EF) = P(F)P(E|F)$ , but it is also equal to  $P(EF) = P(E)P(F|E)$ , since  $P(E|F) = P(EF)/P(F)$  and  $P(F|E) = P(EF)/P(E)$ .

Ex. A box contains 8 blue balls and 4 red balls. We draw two balls from the box without replacement. What is the probability that both are red?

E = first ball red

F = second ball red.

$$P(\text{both balls are red}) = P(EF) = P(E)P(F|E)$$

$$P(E) = 4/12$$

$$P(F|E) = 3/11$$

$$P(EF) = P(E)P(F|E) = 4/12 * 3/11 = 1/11$$

Another way to solve this problem: *(using the methods learned in the previous class)*

$$P(\text{two red}) = \binom{4}{2} / \binom{12}{2} = 4!/(2!2!)/12!/(10!2!) = 1/11$$

Multiplication Rule:

$$P(E_1 E_2 \dots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \dots P(E_n|E_1 E_2 \dots E_{n-1})$$

Proof:

$$\begin{aligned} & P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \dots P(E_n|E_1 E_2 \dots E_{n-1}) \\ &= P(E_1) * (P(E_1 E_2)/P(E_1)) * (P(E_1 E_2 E_3)/P(E_1 E_2)) \dots P(E_1 E_2 \dots E_n)/P(E_1 E_2 \dots E_{n-1}) \\ &= P(E_1 E_2 \dots E_n) \end{aligned}$$

Ex. A box contains five red balls and five green balls. Four balls are sampled without replacement. What is the probability of drawing four red balls?

$E_1$  = first ball is red

$E_2$  = second ball is red

$E_3$  = third ball is red

$E_4$  = fourth ball is red

$$P(E_1 E_2 E_3 E_4) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2)P(E_4|E_1 E_2 E_3)$$

$$P(E_1) = 5/10 = 1/2$$

$$P(E_2|E_1) = 4/9$$

$$P(E_3|E_1 E_2) = 3/8$$

$$P(E_4|E_1 E_2 E_3) = 2/7$$

$$P(E_1 E_2 E_3 E_4) = 1/2 * 4/9 * 3/8 * 2/7 = 24/1008 = 1/42$$

Another way to solve this problem: *(using the methods learned in the previous class)*

$$(5 \text{ } 4)/(10 \text{ } 4) = 5/210 = 1/42$$

Ex. Matching problem (revisited) - Suppose that each of three men at a party throws his hat into the center of the room. The hats are first mixed up and then each man randomly selects a hat. What is the probability that none of the three men selects his own hat?

Let us denote by  $E_i$ ,  $i=1, 2, 3$ , the event that the  $i$ th man selects his own hat.

We shall solve the above by first calculating the complementary probability that at least one man selects his own hat.

$$P(\text{no man selects his hat}) = 1 - P(\text{at least one man selects his own hat}) = 1 - P(E_1 \cup E_2 \cup E_3)$$

To calculate the probability  $P(E_1 \cup E_2 \cup E_3)$ , we need Proposition 4:

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

$$P(E_i) = 1/3 \text{ for } i=1,2,3$$

$$P(E_i E_j) = P(E_i)P(E_j|E_i) \quad i \neq j$$

$$P(E_i) = 1/3 \text{ and } P(E_j|E_i) = 1/2$$

$$P(E_i E_j) = 1/3 * 1/2 = 1/6$$

$$P(E_1 E_2 E_3) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2)$$

$$P(E_1)P(E_2|E_1) = P(E_1 E_2) = 1/6$$

$$P(E_3|E_1 E_2) = 1$$

Now we get,

$$P(E_1 \cup E_2 \cup E_3) = 1/3 + 1/3 + 1/3 - 1/6 - 1/6 - 1/6 + 1/6 = 2/3$$

Hence, the probability that none of the men selects his own hat is  $1 - 2/3 = 1/3$ .

### Prisoner's paradox.

Of three prisoners (A, B and C), only the one secretly pardoned by the governor will escape execution. The governor picked at random and told the warden. The warden refuses to tell A his fate, but agrees to name *another* prisoner who's doomed; he reveals that B was not pardoned. The survivor will thus be A or C. What's the probability that A will be pardoned?

Warden thinks the probability is 1/3. A thinks it is 1/2. Who is correct?

The warden is correct. A has the wrong sample space.

$$S_A = \{AB, n(AC), BC\}$$

But A's question to the warden adds an event, the warden's response.

$$S = \begin{array}{ll} \text{A and B executed - Warden says B} & 1/3 \\ \text{A and C executed - Warden says C} & 1/3 \\ \text{B and C executed - Warden says B} \setminus & 1/6 \\ \text{B and C executed - Warden says C} / & 1/6 \end{array}$$

The first two probabilities are 1/3 because C is safe and the warden can't tell A his fate, so the warden has to say B or C, respectively. If A is safe, the warden has two choices.

$$P(\text{Warden says B}) = 1/3 + 1/6 = 1/2$$

$$P(\text{B and C executed} \mid \text{Warden says B}) = (1/6)/(1/2) = 1/3$$

It is very important to note that conditional probabilities are probabilities in their own right. By this we mean that conditional probability satisfy all the properties of ordinary probabilities.

(1)  $0 \leq P(E|F) \leq 1$

(2)  $P(S|F) = 1$

(3) If  $E_i, i=1,2,\dots$  are mutually exclusive events, then  $P(\cup E_i|F) = \sum P(E_i|F)$

*Easy to show.*

*Still a probability, but the sample space has been redefined.*

All the Propositions we derived for regular probabilities also hold for conditional probabilities.

Ex.  $P(A^C|B) = 1 - P(A|B)$

## **Bayes' Formula**

Let E and F be events. We can express E as  $E = EF \cup EF^C$

Therefore by Axiom 1:  $P(E) = P(EF) + P(EF^C)$

But  $P(EF) = P(E|F)P(F)$  and  $P(EF^C) = P(E|F^C)P(F^C)$

So,  $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$

Another way of writing this is  $P(E) = P(E|F)P(F) + P(E|F^C)(1-P(F))$

This formula enables us to calculate the probability of an event by conditioning upon whether or not a second event occurs.

*Very useful!*

Ex. Box 1 contains two white balls and three blue balls, while Box 2 contains three white and four blue balls. A ball is drawn at random from Box 1 and put into Box 2, and then a ball is picked at random from Box 2 and examined. What is the probability it is blue?

A = ball chosen from Box 1 is blue.

B = ball chosen from Box 2 is blue.

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)$$

$$P(B|A) = 5/8$$

$$P(A) = 3/5$$

$$P(B|A^C) = 4/8$$

$$P(A^C) = 2/5$$

$$P(B) = 5/8 * 3/5 + 4/8 * 2/5 = 23/40$$

An important equation to note is that,

$$P(F|E) = P(EF)/P(E) = P(E|F)P(F)/P(E) = P(E|F)P(F) / [ P(E|F)P(F) + P(E|F^C)P(F^C) ]$$

*Bayes' Formula - will discuss a more general version soon.*



Ex. A lab test is 95% correct in detecting a certain disease when the disease is actually present. However the test also yields a false positive for 1% of the healthy people tested. If 0.5% of the population has the disease, what is the probability that somebody has the disease, given that he tests positive.

A = person has the disease.

B = person tests positive.

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

$$P(B|A) = 0.95$$

$$P(A) = 0.005$$

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C) = 0.95*0.005 + 0.01*0.995 = 0.0147$$

$$P(A|B) = 0.95*0.005/0.0147 = 0.323$$

What if we test twice?

B' = person tests positive twice.

$$P(B'|A) = P(B|A) P(B|A) = 0.95^2$$

$$P(B') = P(B'|A)P(A) + P(B'|A^C)P(A^C) = 0.95*0.95*0.005 + 0.01*0.01*0.995 = 0.0046$$

$$P(A|B') = P(B'|A)P(A)/P(B') = 0.95*0.95*0.005/0.0046 = 0.98.$$

Ex. Morse Code: {A .-}, {B -...}, {C -.-.}, etc....

When Morse code messages are sent, there are sometimes errors in transmission. Assume that the proportion of dots to dashes in a particular message is 3:4. Further assume that with probability 1/8 a dot is received as a dash, and vice versa.

What is the probability a dot is received? What is the probability a dot is sent given that a dot is received?

A = dot sent

B = dot received

$$P(A) = 3/7, P(A^C) = 4/7$$

$$P(B^C|A) = 1/8, P(B|A^C) = 1/8$$

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C) = (7/8)*(3/7) + (1/8)*(4/7) = (25/56)$$

$$P(A|B) = P(B|A)P(A)/P(B) = (7/8)*(3/7)/(25/56) = 21/25$$

Ex. In answering a multiple choice question the student either knows the answer or guesses. Let  $p$  be the probability the student knows the answer and  $(1-p)$  the probability he guesses. Assume that a student who guesses gets the question correct with probability  $1/m$ , where  $m$  are the number of choices. What is the probability that the student knew the answer to the question, given that he gets the correct answer?

$A$  = student knows answer

$B$  = student gets answer correct

$$P(A|B) = P(B|A)P(A)/P(B)$$

$$P(B|A) = 1$$

$$P(A) = p$$

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C) = 1 \cdot p + 1/m \cdot (1-p) = p + (1-p)/m$$

$$P(A|B) = p/(p + (1-p)/m) = mp/(1-p+mp)$$

$$m = 4 \text{ and } p = 1/2$$

$$P(A|B) = 2/(1-1/2+2) = 2/(5/2) = 4/5$$

***Law of total probability.*** (denominator in the Bayes' ratio)

We have so far seen that,  $P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$ .

In general if  $F_1, F_2, \dots, F_n$  are mutually exclusive events such that  $\bigcup F_i = S$  then

$$P(E) = \sum_i P(E|F_i)P(F_i).$$

***Bayes' Formula.***

We have also seen in the previous few examples that  $P(E|F) = P(EF)/P(F) = P(F|E)P(E)/P(F)$ .

A general version of this equation is given by Bayes' Formula.

**Bayes' Formula:** Suppose that  $F_1, F_2, \dots, F_n$  are mutually exclusive events, such that  $S = \bigcup_i F_i$ ; then

$$P(F_i|E) = P(E|F_i)P(F_i)/(\sum_j P(E|F_j)P(F_j)).$$

Ex. There are three machines (A, B and C) at a factory that are used to make a certain product. Machine A makes 25% of the products, B 35% and C 40%. Of the products that A makes, 5% are defective, while for B 4% are defective and for C 2%. The products from the different machines are mixed up and sent to the customer.

- (a) What is the probability that a customer receives a defective product?
- (b) What is the probability that a randomly chosen product was made by A, given the fact that it is defective?

E = product defective

F1 = product comes from A

F2 = product comes from B

F3 = product comes from C

$$(a) P(E) = P(E|F1)P(F1) + P(E|F2)P(F2) + P(E|F3)P(F3) = 0.05*0.25 + 0.04*0.35 + 0.02*0.40 = 0.0345$$

$$(b) P(F1|E) = 0.25*0.5/0.0345 = 0.36$$

Ex. Monty Hall Problem - There are 3 doors, behind one of which is a prize. The host asks you to pick any door. Say you pick door A. The host then opens door B and shows there is nothing behind door B. He then gives you the choice of either sticking with your original choice of door A, or switching to door C. Should you switch?

The *a priori* probability that the prize is behind door X,  $P(X) = 1/3$

Let,

A = prize behind A

B = prize behind B

C = prize behind C

E = Host opens B

The probability that the host opens door B if the prize were behind A,  
 $P(E|A) = 1/2$

The probability that the host opens door B if the prize were behind B,  
 $P(E|B) = 0$

The probability that the host opens door B if the prize were behind C,  
 $P(E|C) = 1$

The probability that Monty opens door B is therefore,  
$$P(E) = P(A)*P(E|A) + P(B)*P(E|B) + P(C)*P(E|C)$$
$$= 1/6 + 0 + 1/3 = 1/2$$

Then, by Bayes' Theorem,

$$P(A|E) = P(A)*P(E|A)/P(E) = (1/6)/(1/2) = 1/3$$

and

$$P(C|E) = P(C)*P(E|C)/P(E) = (1/3)/(1/2) = 2/3$$

In other words, the probability that the prize is behind door C is higher when the host opens door B, and you SHOULD switch!

Assume that  $F_i, i=1, \dots, n$  are competing hypothesis, then Bayes formula gives us a way to compute the conditional probabilities of these hypotheses when additional evidence  $E$  becomes available.

If  $E$  has occurred, what is the probability that  $F_i$  occurred as well. Bayes' Formula gives us a way to update our personal probability.

In the context of Bayes' Formula,  $P(F_i)$  is the *prior probability* of  $F_i$  and  $P(F_i | E)$  is the *posterior distribution* of  $F_i$ .

Ex. A plane is missing and it is assumed that it has crashed in any of three possible regions with equal probability. Assume that the probability of finding the plane in Region 1, if in fact the plane is located in there, is 0.8. What is the probability the plane is in the  $i$ -th region given that the search of Region 1 is unsuccessful?

$R_1$  = plane is in region 1

$R_2$  = plane is in region 2

$R_3$  = plane is in region 3

$E$  = search of region 1 unsuccessful.

$$P(R_1|E) = P(E|R_1)P(R_1)/(\sum P(E|R_i)P(R_i))$$

$$P(R_i) = 1/3$$

$$P(E|R_1) = 0.2 = 1/5$$

$$P(E|R_2) = 1$$

$$P(E|R_3) = 1$$

$$P(R_1|E) = (1/5 * 1/3) / (1/5 * 1/3 + 1/3 + 1/3) = (1/15) / (11/15) = 1/11$$

$$P(R_2|E) = (1 * 1/3) / (1/5 * 1/3 + 1/3 + 1/3) = (1/3) / (11/15) = 5/11$$

$$P(R_3|E) = (1 * 1/3) / (1/5 * 1/3 + 1/3 + 1/3) = (1/3) / (11/15) = 5/11$$

Original hypothesis:  $P(R_1) = P(R_2) = P(R_3)$

New hypothesis:  $P(R_1) < P(R_2) = P(R_3)$

## Odds ratio

The odds ratio of an event A is given by,  $P(A)/P(A^C) = P(A)/(1-P(A))$

The odds ratio of an event A tells us how much more likely it is that an event A occurs than it is that it doesn't occur.

If the odds ratio is equal to alpha, then one says that the odds are alpha to 1 in favor of the hypothesis.

Ex. A box contains 3 red and 1 blue ball.

The probability of drawing a red ball is  $3/4$ .

The odds of picking a red ball is 3 to 1.

A hypothesis H is true with probability  $P(H)$ . The odds ratio is  $P(H)/P(H^C)$ . New evidence E is introduced.

$$P(H|E) = P(E|H)P(H)/P(E) \text{ and } P(H^C|E) = P(E|H^C)P(H^C)/P(E)$$

Odds ratio after introduction of new evidence E is,

$$\begin{aligned} P(H|E)/P(H^C|E) &= (P(E|H)P(H)/P(E)) * (P(E|H^C)P(H^C)/P(E)) = \\ &= (P(H)/P(H^C)) * (P(E|H)/P(E|H^C)). \end{aligned}$$

Odds ratio increases if  $P(E|H) > P(E|H^C)$ .

### **Independent Events.**

*In most of the examples we have studied so far the occurrence of an event A changes the probability of an event B occurring.*

In some cases the occurrence of a particular event, B, has no effect on the probability of another event A.

In this situation we can say,  $P(A|B) = P(A)$ .

Since,  $P(A|B) = P(AB)/P(B)$  we have that  $P(AB) = P(A)*P(B)$

**Definition:** Two events A and B, are statistically independent if  $P(A \cap B) = P(A)*P(B)$

If two events are not independent they are said to be dependent.

Two events are independent if the occurrence of one does not change the probability of the other occurring.

Question: If A and B are mutually exclusive are they independent?

**Ex.** Flip two coins

E = H on first toss

F = H on second toss

$$P(E \cap F) = P(HH) = \frac{1}{4}$$

$$P(E) = \frac{1}{2}, P(F) = \frac{1}{2}$$

$$P(E)P(F) = P(E \cap F)$$

E and F are independent.

Theorem: If E and F are independent, so are

- (a) E and  $F^C$ .
- (b)  $E^C$  and F.
- (c)  $E^C$  and  $F^C$ .

Proof of (a):

Assume E and F are independent,  $P(EF) = P(E)P(F)$

$$E = EF \cup EF^C$$

EF and  $EF^C$  are mutually exclusive, so

$$P(E) = P(EF) + P(EF^C)$$

or

$$P(EF^C) = P(E) - P(EF) = P(E) - P(E)P(F) = P(E)(1 - P(F)) = P(E)P(F^C)$$

Hence the result is proved. Show (b) and (c) at home.

Definition: The three events E, F and G are said to be independent if

$$P(EFG) = P(E)P(F)P(G)$$

$$P(EF) = P(E)P(F)$$

$$P(EG) = P(E)P(G)$$

$$P(FG) = P(F)P(G)$$

Ex. Roll two dice. Let,

A = first die is a 4

B = second die is a 3

C = sum of two dice is 7 =  $\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

$$P(A) = 1/6$$

$$P(B) = 1/6$$

$$P(C) = 1/6$$

$$P(A)P(B) = 1/6 * 1/6 = 1/36$$

$$P(AB) = P(\{4,3\}) = 1/36$$

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

Using similar arguments, we can see that  $P(AC) = P(A)P(C)$  and  $P(BC) = P(B)P(C)$ .

*(Verify for yourself)*

$P(ABC) = P(\{4,3\}) = 1/36$ . But  $P(A)P(B)P(C) = 1/6 * 1/6 * 1/6 = 1/216$ . So A, B, C are NOT independent.



Definition: The events  $E_1, E_2, \dots, E_n$  are independent if, for every subset  $E_{i_1}, E_{i_2}, \dots, E_{i_r}$ ,  $2 \leq r \leq n$ , of these events, we have that  $P(E_{i_1}, E_{i_2}, \dots, E_{i_r}) = P(E_{i_1})P(E_{i_2}) \dots P(E_{i_r})$ .

### Independent Trials

Sometimes an experiment consists of performing a sequence of sub-experiments. If each sub-experiment is identical, then the sub-experiments are called trials.

*If the trials are independent they are called independent trials.*

Ex. Roll a dice five times

What is the probability of rolling a five exactly 2 times?

$E$  = rolling a five

$E^C$  = not rolling a five

$$P(E) = 1/6$$

$$P(E^C) = 5/6$$

$$P(\{(E^C)EE(E^C)(E^C)\}) = 5/6 * 1/6 * 1/6 * 5/6 * 5/6 = (1/6)^2(5/6)^3$$

(5 choose 2) different ways to get exactly 2 fives in five rolls

$$P(\text{Rolling exactly 2 fives in 5 tries}) = (5 \text{ choose } 2)(1/6)^2(5/6)^3$$

What is the probability of at least one 6 in 5 tries?

$$P(1 \text{ six in 5 tries}) = (5 \text{ choose } 1) (1/6)^1(5/6)^4$$

$$P(2 \text{ sixes in 5 tries}) = (5 \text{ choose } 2) (1/6)^2(5/6)^3$$

*Etc....*

*Use complements:*

$$1 - P(0 \text{ sixes in 5 tries}) = 1 - (5 \text{ choose } 0)(1/6)^0(5/6)^5 = 1 - (5/6)^5 = 0.598$$

In general if we have a series of trials where in each trial the event  $E$  consists of a success and the event  $E^C$  consists of a failure, where  $P(E)=p$  and  $P(E^C)=(1-p)$ , then the probability of exactly  $k$  successes in the  $n$  trials is:

$$P(\text{exactly } k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{(n-k)}. \text{ (Binomial probability.)}$$

Theorem: If the events  $E_1, E_2, \dots, E_n$  are independent and  $P(E_i) = p_i$ , the probability that at least one of the events occur is  $1 - (1-p_1)(1-p_2)\dots(1-p_n)$ .

Proof:

Let  $H = \bigcup E_i$ , then  $H^C = (\bigcup E_i)^C = \bigcap E_i^C$  by DeMorgan's Law.

Since the events  $E_1, E_2, \dots, E_n$  are independent, then the events  $E_1^C, E_2^C, \dots, E_n^C$  are also independent.

$$P(H) = 1 - P(H^C) = 1 - P(\bigcap E_i^C) = 1 - P(E_1^C)P(E_2^C) \dots P(E_n^C) = 1 - (1-p_1)(1-p_2)\dots(1-p_n)$$

Corollary: If the events  $E_i$  are independent and each of them occur with a probability  $p$ , the probability that at least one of the events occur is  $1 - (1-p)^n$ .

Ex. A person takes a certain risk at 1000 different opportunities, each time independent of the last. An accident occurs each time with probability  $1/1000$ . What is the probability that an accident will take place at least one of the 1000 opportunities?

$$1 - (1 - 1/1000)^{1000} = 0.63 \approx 1 - \exp(-1).$$

Ex. A series of independent trials are performed, which result in success with probability  $p$  and failure with probability  $1-p$ . What is the probability that exactly  $n$  successes occur before  $m$  failures?

In order for exactly  $n$  successes to occur before  $m$  failures, it is equivalent to that there are exactly  $n$  successes in the first  $m+n-1$  trials. Hence,

$$P = \binom{m+n-1}{n} p^n (1-p)^{(m-1)}.$$

Ex. Independent trials, consisting of rolling a pair of fair dice, are performed. What is the probability that an outcome of 5 appears before an outcome of 7 when the outcome of a roll is the sum of the dice.

Let  $N$  = neither 5 or 7  
 5, N5, NN5, NNN5, .....

Let,  $E_n$  = no 5 or 7 appears on the first  $n-1$  trials and a 5 appears on the  $n$ th trial.

The desired probability is,  $P(UE_n) = \sum_{n=1}^{\infty} P(E_n)$ .

$$P(5) = P(\{(1,4), (2,3), (3,2), (4,1)\}) = 4/36$$

$$P(7) = P(\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}) = 6/36$$

$$P(\text{neither 5 or 7}) = 1 - (4/36 + 6/36) = 26/36 = 13/18$$

$$P(E_n) = (13/18)^{n-1} \cdot 4/36$$

$$\begin{aligned} P(UE_n) &= \sum_{n=1}^{\infty} (13/18)^{n-1} \cdot 4/36 \\ &= 1/9 \sum_{j=0}^{\infty} (13/18)^j \\ &= 1/9 \cdot (1/(1-13/18)) \\ &= 1/9 \cdot 1/(5/18) = 2/5 \end{aligned}$$

We used the formula for the geometric series:

$$\sum_{k=0}^n t^k = (1-t^{n+1})/(1-t) \text{ for } t \neq 1$$

$$\sum_{k=0}^{\infty} t^k = 1/(1-t) \text{ for } |t| < 1$$

To prove this, subtract  $t \cdot \sum t^n$  from  $\sum t^n$ .

Let,  $F_n$  = no 5 or 7 appears on the first  $n-1$  trials and a 7 appears on the  $n$ th trial.

Verify that  $P(UF_n) = 3/5$ .

Ex. A game consists of 5 matches and two players, A and B. Player A wins a match with probability  $2/3$ . Player B wins a match with probability  $1/3$ . The player who first wins three matches is the winner of the game. Assume the matches are independent. What is the probability that A wins the game? Two ways to compute the probability:

I. Player A wins 3 matches and player B wins 2 matches or less. (Game ends when A wins his third match)

A wins three and B two OR A wins three and B one OR A wins three and B zero

*Note A always wins the last game.*

$$P(\text{A wins}) = \binom{4}{2} (2/3)^3 (1/3)^2 + \binom{3}{2} (2/3)^3 (1/3) + \binom{2}{2} (2/3)^3 = 64/81$$

Or: II. Player A wins 3 matches or more out of the 5 matches. (A wins at least 3 matches)

A wins three and B two OR A wins four and B one OR A wins five and B zero

$$P(\text{A wins}) = \binom{5}{3} (2/3)^3 (1/3)^2 + \binom{5}{4} (2/3)^4 (1/3) + \binom{5}{5} (2/3)^5 = 64/81$$

*Both ways are equivalent.*

## Markov chains

I.i.d. sequences are the simplest stochastic process we can imagine: no trial depends on any other. How can we generalize this idea usefully? Let's think about a random walk: flip an i.i.d. coin and keep track of the number of heads. Let  $S(N)$  = number of heads after  $N$  trials. To compute  $p[S(N)=x]$ , we just have to know  $S(N-1)$ ; note that we don't have to keep track of  $S(N-2)$ ,  $S(N-3)$ , etc. This is extremely useful, because it leads to recursive methods for computing  $p[S(N)=x]$  (and related quantities).

The essential feature here is the "Markov property":

$$p[S(N)=x | S(N-1), S(N-2), S(N-3), \dots S(1)] = p[S(N)=x | S(N-1)];$$

if we know the previous state, we can forget the rest of the past.

A random sequence that has this Markov property is called a "Markov chain." These are incredibly useful in a wide variety of applications, for two reasons: first, a great number of real-world processes can be modeled as Markov chains (just think of all the places where random walks pop up – but remember, Markov chains are slightly more general than random walks), and second, we can perform many computations very efficiently in Markov chains, due to their "forgetting" property (this allows us to summarize long, complicated historical sequences  $S(1), S(2), \dots S(N-1)$  with a much simpler sufficient statistic,  $S(N-1)$ ).

To demonstrate how easy some of these computations are, consider a special case: a Markov chain  $X(n)$  with a finite state space, with stationary transition matrix. In this case, we can fully specify the chain by fixing two quantities:

$P(X(1)=i) = q_i$ : the initial conditions

$P(X(n)=i | X(n-1)=j) = P_{ij}$ : the transition matrix.

Now if you hand me a sequence  $X(1), X(2), \dots X(N)$ , I can use Bayes' rule to compute its probability:

$$\begin{aligned} p[X(1), \dots X(N)] &= p[X(1)] p[X(2)|X(1)] p[X(3)|X(1), X(2)] \dots p[X(N)|X(1), \dots X(N-1)] \\ &= p[X(1)] p[X(2)|X(1)] p[X(3)|X(2)] \dots p[X(N)|X(N-1)], \text{ by Markov} \end{aligned}$$

And if you want to compute the probability  $p[X(N)=i]$ , we just have to marginalize: sum the above over  $X(1), X(2), \dots X(N-1)$ :

$$\begin{aligned} \text{Sum}_{\{X(1), X(2), \dots X(N-1)\}} p[X(1), \dots X(N)] \\ = \text{Sum}_{\{X(1), X(2), \dots X(N-1)\}} p[X(1)] p[X(2)|X(1)] p[X(3)|X(2)] \dots p[X(N)|X(N-1)] \end{aligned}$$

We rearrange this:

$$\begin{aligned} = \text{Sum}_{\{X(N-1)\}} p[X(N)|X(N-1)] \text{Sum}_{\{X(N-2)\}} p[X(N-1)|X(N-2)] \dots \\ \text{Sum}_{\{X(1)\}} p[X(2)|X(1)] p[X(1)]. \end{aligned}$$

This can be computed recursively:

$$\text{Sum}_{\{X(1)\}} p[X(2)|X(1)] p[X(1)] = p[X(2)], \text{ then}$$

$$\text{Sum}_{\{X(2)\}} p[X(3)|X(2)] p[X(2)] = p[X(3)], \text{ etc.}$$

In more detail,

$$p[X(2)=j] = \text{Sum}_i p[X(2)=j | X(1)=i] p[X(1)=i] = \text{Sum}_i P_{ij} q_i.$$

This is just a matrix multiplication: the vector  $p[X(2)] = P * q$ .

If we write out the next sum, over  $X(2)$ , we get the same thing:  $p[X(3)] = P * p[X(2)] = P * P * q$ .

More generally,  $p[X(n)] = P^{n-1} * q$ .

This in turn means that we can understand the long-term behavior of  $p[X(n)]$  by understanding the eigenvectors of  $P$ . If  $P$  is diagonalizable,  $P = O D O'$ , then  $p[X(n)] = P^{n-1} * q = O D^{n-1} O' q$ .

Perron-Frobenius theorem guarantees that the largest eigenvalue is one (though the largest eigenvalue might not be unique). If all  $P_{ij} > 0$  (more generally, aperiodic and irreducible), then the largest eigenvalue is unique, and  $p[X(n)]$  tends to the largest eigenvector of  $P$ .