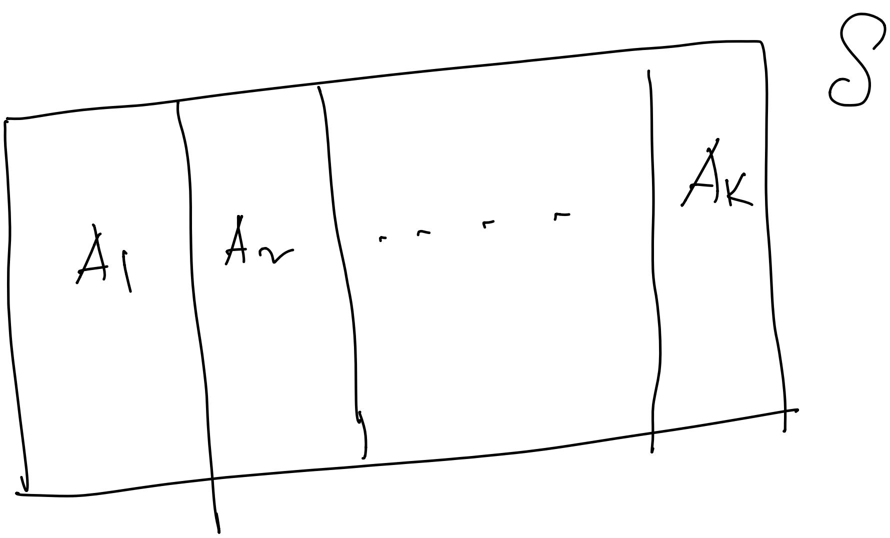
Partitioning of the sample space:

Events $A_1, A_2, \ldots, A_K$ are said to partition the sample space S if:

1. $A_i \cap A_j = \emptyset$ for $i \neq j$

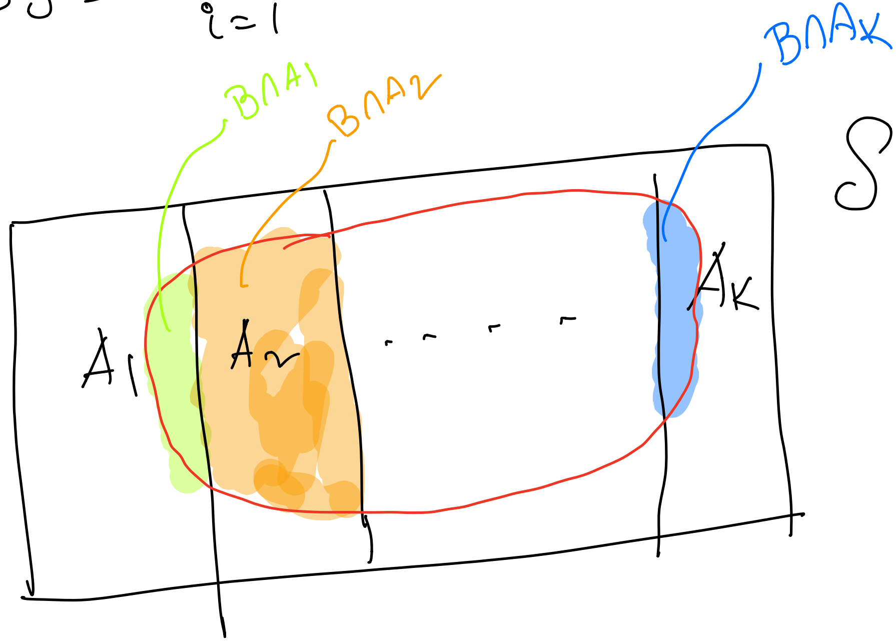
2. $\bigcup_{i=1}^{K} A_i = S$



Law of total probability:

Suppose $A_1, \ldots A_K$ forms a partition of the sample space. Then for any event B in the sample space, we have

$$P[B] = \sum_{i=1}^{K} P[B|A_i]P[A_i]$$



Bayes Rule:

Let $\{A_1, A_2, \ldots A_K\}$ be a partition of the sample space S, and suppose each of the events $A_1, A_2, \ldots A_K$ has nonzero probability. Let B be any event for which $P(B) > 0$. Then for each integer n $(1 \leq n \leq K)$, we have Bayes formula:

$$P(A_n | B) = \frac{P(A_n) P(B | A_n)}{\sum_{j=1}^{K} P(A_j) P(B | A_j)}$$

Chain Rule:

For any events A and B, we have

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

more generally, for any events $A_1, \dots A_n$,

$$P\left(\cap_{i=1}^{n} A_i\right) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1,A_2)\cdots P\left(A_n|\cap_{i=1}^{n-1} A_i\right)$$

To prove the above, we will use induction on n. The base case is $n=1$. For the base case,

$$P(A_1) = P(A_1)$$

which is trivially true. For the inductive step, let $n>1$ and assume (the inductive hypothesis) that

$$P\left(\cap_{i=1}^{n-1} A_i\right) = P(A_1)\,P(A_2|A_1)\cdots P\left(A_{n-1}|\cap_{i=1}^{n-2} A_i\right)$$

Then,

$$P(\cap_{i=1}^{n} A_i) = P\left(A_n \cap \{\cap_{i=1}^{n-1} A_i\}\right)$$

By the definition of conditional probability,

$$P(\cap_{i=1}^{n} A_i) = P(A_n \mid \cap_{i=1}^{n-1} A_i)\, P(\cap_{i=1}^{n-1} A_i)$$

Now, by the induction hypothesis,

$$P(\cap_{i=1}^{n} A_i) = P(A_n \mid \cap_{i=1}^{n-1} A_i)\, P(A_1) P(A_2 \mid A_1) \cdots P(A_{n-1} \mid \cap_{i=1}^{n-2} A_i)$$

This completes the proof by induction.

Practice Problems on probability basics:

1/ Let's define the following events:

Z_D: A randomly selected chip is defective

Z_A: A randomly selected chip was manufactured by A.

Z_B: A randomly selected chip was manufactured by B.

Z_C: A randomly selected chip was manufactured by C.

Now, we want to compute

$P(Z_A | Z_D)$, $P(Z_C | Z_D)$

from the problem statement, we know

$P(Z_D|Z_A) = 0.005$, $P(Z_D|Z_B) = 0.001$

$P(Z_D|Z_C) = 0.01$.

By chain rule of probability,

$$P(Z_A|Z_D) = \frac{P(Z_D|Z_A)\, P(Z_A)}{P(Z_D)}$$

Since Z_A, Z_B and Z_C forms a partition of the sample space, so by law of total probability

$$P(Z_D) = P(Z_D|Z_A)P(Z_A) + P(Z_D|Z_B)P(Z_B) + P(Z_D)Z_C)\, P(Z_C)$$

$$= 0.005 \times 0.5 + 0.001 \times 0.1 + 0.01 \times 0.4$$

$$= 0.0025 + 0.0001 + 0.004$$

$$P(Z_D) = 0.0066$$

Hence,

$$P(Z_A|Z_D) = \frac{0.005 \times 0.5}{0.0066} = 0.3788$$

$$P(Z_C|Z_D) = \frac{0.01 \times 0.4}{0.0066} = 0.606$$

2

Suppose we define the following events:

T: A two headed coin is flipped

F: A fair coin is flipped

B: A biased coin is flipped

H: The flipped coin shows a head

a) Since T, F, and B forms a partition of the sample space, so by law of total probability

$$P(H) = P(H|T)P(T) + P(H|F) \cdot P(F) + P(H|B)P(B)$$

$= P(T) + \frac{1}{2} P(F) + p P(B)$

$= \frac{1}{3} + \frac{1}{6} + \frac{1}{3} p$

$\therefore \ P(H) = \frac{1}{2} + \frac{1}{3} p$

Since $0 \le p \le 1$, so

$\max P(H) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$

b) By the definition of conditional probability

$$P(F|H) = \frac{P(H|F) \cdot P(F)}{P(H)}$$

$$P(F|H) = \frac{1/2 \cdot 1/3}{1/2 + 1/3 p} = \frac{1/6}{1/2 + 1/3 p}$$

Since $0 \le p \le 1$, so $P(F|H)$ is maximized when $p=0$

$$\max P(F|H) = 1/3$$

$P(F|H)$ is minimized when $p=1$

$$\min P(F|H) = \frac{1/6}{5/6} = 1/5$$

3

Let's define the following events:

D: man has a dangerous type of the disease

T: man has a positive PSA test

From the problem statement, we are given the following quantities

$$P(T|D) = 0.9$$

$$P(T|D^c) = 0.01$$

$$P(D) = 0.0005$$

a) By bayes law,

$$P(D|T) = \frac{P(T|D)\,P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)}$$

$$= \frac{(0.9)(0.0005)}{(0.9)(0.0005) + (0.01)(0.9995)}$$

$$= \frac{0.00045}{0.00045 + 0.009995}$$

$$= \frac{0.00045}{0.010445}$$

$$P(D|T) = 0.043$$

b) Again by bayes law,

$$P(D|T^c) = \frac{P(T^c|D)\,P(D)}{P(T^c)}$$

$$= \frac{(0.1)(0.0005)}{(0.1)(0.0005) + (0.99)(0.9995)}$$

$$= \frac{0.00005}{0.00005 + 0.989505}$$

$$= \frac{0.00005}{0.989555}$$

$$P(D|T^c) = 0.000050528$$

5

a) Suppose we define X_i as the bernoulli random variable

$$X_i = \begin{cases} 1, & i\underline{th} \text{ plate is isolated} \\ 0, & i\underline{th} \text{ plate is not isolated} \end{cases}$$

Then,

$$X = \sum_{i=1}^{15} X_i$$

Now,

$$E[X_i] = P(X_i = 1)$$

Then let's compute $P(X_i = 1)$.

$P(X_i = 1)$

$= P(i^{\text{th}}$ plate is isolated$)$

If we define the following events:

B: i^{th} plate is blue

R: i^{th} plate is red

G: i^{th} plate is green

Since B, R, G partitions the sample space so by law of total probability,

$P(i^{\text{th}}$ plate is isolated$)$

$$= P(i \mid B)P(B) + P(i \mid R)P(R) + P(i \mid G)P(G)$$

$$= \left(\frac{10}{14}\right) \cdot \left(\frac{9}{13}\right) \frac{1}{3} \cdot 3 = \frac{45}{91}$$

Hence by linearity of expectation,

$$E[X] = 15\, E[X_i] = 15 \cdot \frac{45}{91} = 7.42$$

b) Suppose we define Y_i as the bernoulli random variable

$$Y_i = \begin{cases} 1, & i^{th} \text{ plate is semi-happy} \\ 0, & i^{th} \text{ plate is not semi-happy} \end{cases}$$

Then,

$$Y = \sum_{i=1}^{15} Y_i$$

Now,

$$E[Y_i] = P(Y_i = 1)$$

Then let's compute $P(Y_i = 1)$.

$P(Y_i = 1)$

$= P(i^{th}$ plate is semi-happy$)$

If we define the following events:

B: i^{th} plate is blue

R: i^{th} plate is red

G: i^{th} plate is green

Since B, R, G partitions the sample space so by law of total probability,

$P(i^{th}$ plate is semi-happy$)$

$= P(i \mid B)P(B) + P(i \mid R)P(R)$
$\quad + P(i \mid G)P(G)$

$$= \left[\frac{10}{14} \cdot \frac{4}{13} + \frac{4}{14} \cdot \frac{10}{13}\right] \cdot \frac{1}{3} \cdot 3$$

$$= \frac{40}{91}$$

Hence by linearity of expectation,

$$E[Y] = 15\,E[Y_i] = \frac{15 \cdot 40}{91} = 6.593$$

c) Suppose we define Z_i as the bernoulli random variable

$$Z_i = \begin{cases} 1, & i^{th} \text{ plate is semi-happy} \\ 0, & i^{th} \text{ plate is not semi-happy} \end{cases}$$

Then,

$$Z = \sum_{i=1}^{15} Z_i$$

Now,

$$E[Z_i] = P(Z_i = 1)$$

Then let's compute $P(Z_i = 1)$.

$P(z_i = 1)$

$= P(i^{\underline{th}}$ plate is joyous $)$

If we define the following events:

B: $i^{\underline{th}}$ plate is blue

R: $i^{\underline{th}}$ plate is red

G: $i^{\underline{th}}$ plate is green

Since B, R, G partitions the sample space so by law of total probability,

$P(i^{\underline{th}}$ plate is joyous $)$

$$= P(i \mid B)P(B) + P(i \mid R)P(R) + P(i \mid G)P(G)$$

$$= \left[\frac{4}{14} \cdot \frac{3}{13}\right] \cdot \frac{1}{3} \cdot 3$$

$$= 6/91$$

Hence by linearity of expectation,

$$E[Z] = 15\,E[Z_i] = 15 \cdot \frac{6}{91} = 0.989$$

4

Since $n = 1, 2, 3, 4, \ldots\ldots$, where n denotes the round in which the duel ends, partitions the sample space so by law of total probability:

a) $P(\text{Jack not hit})$

$$= \sum_{n=1}^{\infty} P(\text{Jack not hit}, n)$$

Now, if duel ends in n rounds and Jack is not hit then,

$$\underset{1}{\underline{Bm}} \quad \underset{2}{\underline{Bm}} \quad \underset{3}{\underline{Bm}} \quad \underset{4}{\underline{BM}} \quad \cdots\cdots \quad \underset{n-1}{\underline{BM}} \quad \underset{n}{\underline{Jill}}$$

So,

P(Jack not hit, n)

$$= (1-P_1)^{(n-1)}(1-P_2)^{(n-1)} \cdot P_1(1-P_2)$$

$$= P_1(1-P_1)^{(n-1)}(1-P_2)^{n}$$

So,

$P(\text{Jack not hit})$

$$= p_1 \sum_{n=1}^{\infty} (1-p_1)^{(n-1)} (1-p_2)^n$$

Since $\displaystyle\sum_{n=1}^{\infty} (1-p_1)^{(n-1)} (1-p_2)^n$

is sum of a geometric series with

$a = 1-p_2$, $r = (1-p_1)(1-p_2)$

So

$P(\text{Jack not hit})$

$$= \frac{p_1 (1-p_2)}{1-(1-p_1)(1-p_2)}$$

b) following similar steps as a) we have

P(both duelists are hit)

$$= \sum_{n=1}^{\infty} (1-p_1)^{(n-1)} (1-p_2)^{(n-1)} p_1 p_2$$

$$= p_1 p_2 \sum_{n=1}^{\infty} (1-p_1)^{(n-1)} (1-p_2)^{(n-1)}$$

$$= \frac{p_1 p_2}{1-(1-p_1)(1-p_2)}$$

c) Since duel can end after $n^{\underline{th}}$ round of shots in 3 possible ways:

→ Jack hit

→ Jill hit

→ Both hit

Then, by law of total probability

P(duel ends after $n^{\underline{th}}$ round)

$$= (1-p_1)^{(n-1)}(1-p_2)^{(n-1)}(1-p_1)p_2$$

$$+ (1-p_1)^{(n-1)}(1-p_2)^{(n-1)}p_1(1-p_2)$$

$$+ (1-p_1)^{(n-1)}(1-p_2)^{(n-1)}p_1 p_2$$

$$= \left[(1-p_1)(1-p_2) \right]^{(n-1)} \left[1 - (1-p_1)(1-p_2) \right]$$

Practice problem on linear algebra basics

(a) Let λ_i be an eigenvalue of A with corresponding eigenvector v_i. Then,

$$A v_i = \lambda_i v_i$$

multiplying both sides on the left by v_i^T. we get,

$$v_i^T A v_i = \lambda_i v_i^T v_i$$

$$\Rightarrow v_i^T A v_i = \lambda_i \|v_i\|_2^2$$

Since $b^T A b > 0$ for all $b \in \mathbb{R}^n$, so

$$v_i^T A v_i = \lambda_i \|v_i\|_2^2 > 0$$

Since $\|v_i\|_2^2 > 0$, so $\lambda_i > 0$.

$\therefore$ All eigenvalues of A are positive.

(b) Let λ_i be an eigenvalue of A with corresponding eigenvector v_i. Then,

$$A v_i = \lambda_i v_i$$

multiplying both sides to the left by A^T, we get

$$A^T A v_i = \lambda_i A^T v_i$$

Since A is an orthogonal matrix, so

$$A^T A = A A^T = I$$

Hence,

$$v_i = \lambda_i A^T v_i$$

Taking the 2-norm of both sides

$$\|v_i\|_2^2 = |\lambda_i|^2 \|A^T v_i\|_2^2$$

Now,

$$\|A^T v_i\|_2^2 = (A^T v_i)^T (A^T v_i)$$

$$= v_i^T A A^T v_i = \|v_i\|_2^2$$

Hence, we have

$$\|v_i\|_2^2 = |\lambda_i|^2 \|v_i\|_2^2$$

$\therefore$ $|\lambda_i| = 1$. Hence, all eigenvalues of A have norm 1.

ⓒ $A \in \mathbb{R}^{m \times n}$ has the following SVD

$$A = U \Sigma V^T$$

Now,

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T$$

$$= U\Sigma V^T V \Sigma^T U^T$$

$$AA^T = U \Sigma \Sigma^T U^T$$

Now, $\Sigma = \begin{bmatrix} \sigma_1 & \sigma_2 & \cdots & \sigma_r & 0 & \cdots & 0 \end{bmatrix}$

So, $\Sigma\Sigma^T = \begin{bmatrix} \sigma_1^2 & & & & & \\ & \sigma_2^2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r^2 & & \\ & & & & 0 & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$

since $\Sigma\Sigma^T$ is a diagonal matrix and U is an orthogonal matrix so the eigendecomposition of the symmetric matrix AA^T is

$$U \Sigma \Sigma^T U^T$$

Hence,

$$\lambda_i(AA^T) = \sigma_i^2(A)$$

2.

a) $(A^T)^T A^T$

$= AA^T$

$= I$ (Since A is orthogonal)

Also,

$A^T (A^T)^T$

$= A^T A$

$= I$ (Since A is orthogonal)

Hence, A^T is also orthogonal

b) $(AB)^T(AB)$

$= B^T A^T A B$

$= B^T I B$ (since A is orthogonal)

$= B^T B$

$= I$ (since B is orthogonal)

Also, $AB(AB)^T$

$= AB B^T A^T$

$= A I A^T$ (Since B is orthogonal)

$= A A^T$

$= I$ (since A is orthogonal)

$\therefore$ AB is orthogonal

c) Let $A = B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Then A and B are both orthogonal,

but

$$A+B = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

is not orthogonal.

d) Suppose the column vectors of A are orthonormal. Hence

$$a_i^T a_i = 1 \quad \forall i$$

$$a_i^T a_j = 0 \quad i \neq j$$

which implies

$$A^T A = I$$

Since $A^{-1} = A^T$, so A is an orthogonal matrix. From (a), we know that A^T is also orthogonal. Since A^T is also orthogonal, so

$$(A^T)^T A^T = I$$

The above relation implies A^T has orthonormal columns, meaning that

A has orthormal rows.