Conditional Probability and Independence

Ruben Zamar
Department of Statistics
UBC

January 22, 2019

MOTIVATION

- The outcome could be any element in the Sample Space, Ω.
- Sometimes the range of possibilities is restricted because of "partial information"
- Examples
  - number of shots:

    partial info: we know it wasn't an "ace"
  - ELEC 321 final grade:

    partial info: we know it is at least a "B"

CONDITIONING EVENT

- The event B representing the "partial information" is called "conditioning event"
- Denote by A the event of interest
- Example (Number of Shots)

$$B = \{2, 3, ...\} = \{\text{not an "ace"}\} \quad \text{(conditioning event)}$$

$$A = \{1, 3, 5, ...\} = \{\text{server wins}\} \quad \text{(event of interest)}$$

- Example (Final Grade)

$$B = [70, 100] = \{\text{at least a "B"}\} \quad \text{(conditioning event)}$$

$$A = [80, 100] = \{\text{an "A"}\} \quad \text{(event of interest)}$$

DEFINITION OF CONDITIONAL PROBABILITY

- Suppose that $P(B) > 0$

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

- The left hand side is read as **"probability of A given B"**

- Useful formulas:

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

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

- $P(A|B)$, as a function of A (and for B fixed) satisfies all the probability axioms:

  - $P(\Omega|B) = P(\Omega \cap B) / P(B) = P(B) / P(B) = 1$

  - $P(A|B) \geq 0$

  - If $\{A_i\}$ are disjoint then

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

$$= \frac{P[\cup (A_i \cap B)]}{P(B)}$$

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

EXAMPLE: NUMBER OF SHOTS

- For simplicity, suppose that points are decided in at most 8 shots, with probabilities:

Shots	1	2	3	4	5	6	7	8
Prob.	0.05	0.05	0.15	0.10	0.20	0.10	0.20	0.15

- Using the table above:

$$\begin{aligned} P(\text{Sever wins} \mid \text{Not an ace}) &= \frac{P(\{3,5,7\})}{P(\{2,3,4,5,6,7,8\})} \\ &= \frac{0.55}{0.95} \\ &= 0.579 \end{aligned}$$

EXAMPLE: FINAL GRADE

- Suppose that

$$P\,(\text{Grade is larger than } x) = \frac{100 - x}{100} = 1 - \frac{x}{100}$$

- Using the formula above:

$$\begin{aligned} P\,(\text{To get an "A"} \mid \text{To get at least a "B"}) &= \frac{P\,([80, 100])}{P\,([70, 100])} \\ &= \frac{100 - 80}{100 - 70} = \frac{20}{30} \\ &= 0.667 \end{aligned}$$

SCREENING TESTS

- Items are submitted to a screening test before shipment
- The screening test can result in either

POSITIVE (indicating that the item may have a defect)

NEGATIVE (indicating that the item doesn't have a defect)

- Screening tests face two types of errors

FALSE POSITIVE

FALSE NEGATIVE

SCREENING TESTS (continued)

- For each item we have 4 possible events

  Item true status:

  $D = \{\text{item is defective}\}$

  $D^c = \{\text{item is not defective}\}$

  Test result:

  $B = \{\text{test is positive}\}$

  $B^c = \{\text{test is negative}\}$

- The following conditional probabilities are normally known

$$\text{Sensitivity of the test:} \quad P(B|D) = 0.95 \ \ \text{(say)}$$

$$\text{Specificity of the test:} \quad P(B^c|D^c) = 0.99 \ \ \text{(say)}$$

which implies

$$P(B^c|D) = 0.05 \quad \text{and} \quad P(B|D^c) = 0.01$$

- The proportion of defective items is also normally known

$$P(D) = 0.02 \ \ \text{(say)}$$

TEST PERFORMANCE

- The following questions may be of interest:
  - What is the probability that a randomly chosen item tests positive?
  - What is the probability of defective given that the test resulted negative?
  - What is the probability of defective given that the test resulted positive?
  - What is the probability of screening error?
- We will compute these probabilities

PROBABILITY OF TESTING POSITIVE

$$
\begin{aligned}
P(B) &= P(B \cap D) + P(B \cap D^c) \\
&= P(D)\,P(B|D) + P(D^c)\,P(B|D^c) \\
&= 0.02 \times 0.95 + (1 - 0.02) \times 0.01 \\
&= 0.0288
\end{aligned}
$$

PROB OF DEFECTIVE GIVEN A POSITIVE TEST

$$
\begin{aligned}
P(D|B) &= \frac{P(D \cap B)}{P(B)} \\
&= \frac{P(D)\,P(B|D)}{P(B)} \\
&= \frac{0.02 \times 0.95}{0.0288} \\
&= 0.65972
\end{aligned}
$$

PROB OF DEFECTIVE GIVEN A NEGATIVE TEST

$$
\begin{aligned}
P\left(D \middle| B^c\right) &= \frac{P\left(D \cap B^c\right)}{P\left(B^c\right)} \\
&= \frac{P\left(D\right) P\left(B^c \middle| D\right)}{1 - P\left(B\right)} \\
&= \frac{0.0098}{1 - 0.0288} \\
&= 0.01
\end{aligned}
$$

SCREENING ERROR

$$
\begin{aligned}
P(\text{Error}) &= P(D \cap B^c) + P(D^c \cap B) \\
&= P(D)\, P(B^c | D) + P(D^c)\, P(B | D^c) \\
&= 0.02 \times (1 - 0.95) + (1 - 0.02) \times 0.01 \\
&= 0.0108
\end{aligned}
$$

BAYES' FORMULA

- The formula

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

  is the simple form of Bayes' formula.

- This has been used in the "Screening Example" presented before.

BAYES' FORMULA (continued)

- The general form of Bayes' Formula is given by

$$P(D_i|B) = \frac{P(D_i \cap B)}{P(B)} = \frac{P(B|D_i)\,P(D_i)}{\sum_{j=1}^{k} P(B|D_j)\,P(D_j)}$$

where D_1, D_2, ..., D_k is a partition of the sample space Ω:

$$\Omega = D_1 \cup D_2 \cup \cdots \cup D_k$$

$$D_i \cap D_j = \phi, \quad \text{for } i \neq j$$

EXAMPLE: THREE PRISONERS

- Prisoners A, B and C are to be executed
- The governor has selected one of them at random to be pardoned
- The warden knows who is pardoned, but is not allowed to tell
- Prisoner A begs the warden to let him know **which one of the other two prisoners is not pardoned**

- Prisoner A tells the warden: "Since I already know that one of the other two prisioners is not pardoned, you could just tell me who is that"

- Prisoner A adds: "If B is pardoned, you could give me C's name. If C is pardoned, you could give me B's name. And if I'm pardoned, you could flip a coin to decide whether to name B or C."

The warden is convinced by prisoner A's arguments and tells him: "B is not pardoned"

Result: Given the information provided by the Warden, C is now twice more likely to be pardoned than A!

Why? Check the derivations below:

NOTATION:

$$A = \{\text{A is pardoned}\}$$

$$B = \{\text{B is pardoned}\}$$

$$C = \{\text{C is pardoned}\}$$

$$b = \{\text{The warden says ``B is not pardoned''}\}$$

Clearly

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

$$\begin{aligned} P(b|B) &= 0 \quad \text{(warden never lies)} \\ P(b|A) &= 1/2 \quad \text{(warden flips a coin)} \\ P(b|C) &= 1 \quad \text{(warden cannot name A)} \end{aligned}$$

By the Bayes' formula:

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

$$= \frac{\frac{1}{2}\times\frac{1}{3}}{\frac{1}{2}\times\frac{1}{3} + 0\times\frac{1}{3} + 1\times\frac{1}{3}} = \frac{1}{3}$$

Hence

$$P(C|b) = 1 - P(A|b) = 1 - \frac{1}{3} = \frac{2}{3}$$

SCREENING EXAMPLE II

- The tested items have two components: "c_1" and "c_2"
- Suppose

$$
\begin{aligned}
D_1 &= \{\text{Only component “}c_1\text{” is defective}\}, \quad P(D_1) = 0.01 \\
D_2 &= \{\text{Only component “}c_2\text{” is defective}\}, \quad P(D_2) = 0.008 \\
D_3 &= \{\text{Both components are defective}\}, \quad P(D_3) = 0.002 \\
D_4 &= \{\text{Both components are non defective}\}, \quad P(D_4) = 0.98
\end{aligned}
$$

- Let

$$B = \{\text{Screening test is positive}\}$$

- Suppose

$$P(B|D_1) = 0.95$$

$$P(B|D_2) = 0.96$$

$$P(B|D_3) = 0.99$$

$$P(B|D_4) = 0.01$$

SOME QUESTIONS OF INTEREST

- The following questions may be of interest:
  - What is the probability of testing positive?
  - What is the probability that component "c_i" $(i = 1, 2)$ is defective when the test resulted positive?
  - What is the probability that the item is defective when the test resulted negative?
  - What is the probability both components are defective when the test resulted positive?
  - What is the probability of testing error?
- We will compute these probabilities

PROB OF TESTING POSITIVE

$$
\begin{aligned}
P(B) &= P(B \cap D_1) + P(B \cap D_2) + P(B \cap D_3) + P(B \cap D_4) \\
&= 0.01 \times 0.95 + 0.008 \times 0.96 + 0.002 \times 0.99 + 0.98 \times 0.01 \\
&= 0.02896
\end{aligned}
$$

Notice that the probability of defective is

$$P(D) = 0.01 + 0.008 + 0.002 = 0.02$$

PROB OF TESTING NEGATIVE

$$
\begin{aligned}
P(B^c) &= P(B^c \cap D_1) + P(B^c \cap D_2) + P(B^c \cap D_3) + P(B^c \cap D_4) \\
&= 0.01 \times 0.05 + 0.008 \times 0.04 + 0.002 \times 0.01 + 0.98 \times 0.99 \\
&= 0.97104
\end{aligned}
$$

Naturally,

$$P(B) + P(B^c) = 0.02896 + 0.97104 = 1$$

TEST RESULTED POSITIVE

The posterior probabilities given this "data" are:

$$P(D_1|B) = \frac{0.01 \times 0.95}{0.02896} = 0.32804$$

$$P(D_2|B) = \frac{0.008 \times 0.96}{0.02896} = 0.26519$$

$$P(D_3|B) = \frac{0.002 \times 0.99}{0.02896} = 0.06837$$

$$P(D_4|B) = \frac{0.98 \times 0.01}{0.02896} = 0.338\,40$$

$$P(\text{defective}|B) = 1 - P(D_4|B) = 1 - 0.338\,40 = 0.6616$$

TEST RESULTED NEGATIVE

The posterior probabilities given this "data" are:

$$
\begin{aligned}
P(D_1|B^c) &= \frac{0.01 \times 0.05}{0.97104} = 0.00051491 \\
P(D_2|B^c) &= \frac{0.008 \times 0.04}{0.97104} = 0.00032954 \\
P(D_3|B^c) &= \frac{0.002 \times 0.01}{0.97104} = 0.000020596 \\
P(D_4|B^c) &= \frac{0.98 \times 0.99}{0.97104} = 0.99913 \\
P(\text{defective}|B^c) &= 1 - P(D_4|B^c) = 1 - 0.99913 = 0.00087
\end{aligned}
$$

CONCLUSION

Prior reliability of items being sold:

$$P(\text{Defective}) = P(D_1) + P(D_2) + P(D_3) = 0.01 + 0.008 + 0.002 = 0.02$$

Posterior reliability of items being sold:

$$P(\text{Defective}|B^c) = 1 - P(D_4|B^c) = 1 - 0.99913 = 0.00087 < 0.001$$

COST - BENEFIT ANALYSIS

Cost: possibly discarding a small percentage of non-defective items

$$\begin{aligned} P\left(\text{“Testing Positive”} \cap \text{“Non-Defective”}\right) &= P\left(B \cap D_4\right) \\ &= P\left(D_4 | B\right) P\left(B\right) \\ &= 0.33840 \times 0.02896 < 0.01 \end{aligned}$$

Benefit: Relative reliability improvement:

$$\begin{aligned} P\left(\text{Defective}\right) - P\left(\text{Defective} | B^c\right) &= 0.02 - 0.00087 \\ &= 0.019\,13 \end{aligned}$$

Notice that

$$0.02/0.00087 > 22$$

INDEPENDENCE

- **DEFINITION:** Events A and B are independent if

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

- If A and B are independent then

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

and

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

DISCUSSION

- If $P(A) = 1$, then A is independent of all B.

$$P(A \cap B) = P(A \cap B) + \overbrace{P(A^c \cap B)}^{=0} = P(B)$$

$$P(A \cap B) = \overbrace{P(A)}^{=1} P(B)$$

DISCUSSION (Cont)

- Suppose that A and B are non-trivial events ($0 < P(A) < 1$ and $0 < P(B) < 1$)

- If A and B are mutually exclusive ($A \cap B = \phi$) then they cannot be independent because

$$P(A|B) = 0 < P(A)$$

- If $A \subset B$ then they cannot be independent because

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

DISCUSSION (Cont)

- Suppose $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and the numbers are equally likely.
- $A = \{1, 2, 3, 4, 5\}$ and $B = \{2, 4, 6, 8\}$
- $P(A \cap B) = P(\{2, 4\}) = 0.20, \quad P(A) P(B) = 0.5 \times 0.4 = 0.20$
- Hence, A and B are independent
- In terms of probabilities A is half of Ω. On the other hand $A \cap B$ is half of B.

DISCUSSION (continued)

- What happens if

$$P(i) = \frac{i}{55} \quad ?$$

- $P(A \cap B) = P(\{2,4\}) = 6/55 = 0.10909$
  $P(A)\,P(B) = (15/55) \times (20/55) = 0.099174$

- Hence, A and B are not independent in this case.

Definition: We say that the events $A_1, A_2, ..., A_n$ are independent if

$$P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = P\left(A_{i_1}\right) P\left(A_{i_2}\right) \cdots P\left(A_{i_k}\right)$$

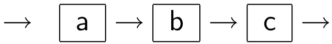
for all $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, and all $1 \leq k \leq n$.

For example, if $n = 3$, then

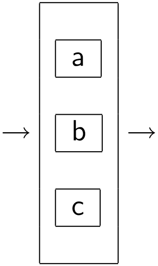
$$
\begin{aligned}
P(A_1 \cap A_2) &= P(A_1)\, P(A_2) \\
P(A_1 \cap A_3) &= P(A_1)\, P(A_3) \\
P(A_2 \cap A_3) &= P(A_2)\, P(A_3) \\
P(A_1 \cap A_2 \cap A_3) &= P(A_1)\, P(A_2)\, P(A_3)
\end{aligned}
$$

SYSTEM OF INDEPENDENT COMPONENTS

- In series



- In parallel



NOTATION

$$A = \{\text{Component } a \text{ works}\}$$

$$B = \{\text{Component } b \text{ works}\}$$

$$C = \{\text{Component } c \text{ works}\}$$

INDEPENDENT COMPONENTS

We assume that A, B and C are independent, that is

$$\begin{aligned}
P(A \cap B \cap C) &= P(A)\,P(B)\,P(C) \\
P(A \cap B) &= P(A)\,P(B), \\
P(B \cap C) &= P(B)\,P(C), \\
P(A \cap C) &= P(A)\,P(C)
\end{aligned}$$

RELIABILITY CALCULATION

Problem 1: Suppose that

$$P(A) = P(B) = P(C) = 0.95.$$

Calculate the reliability of the system

$$\longrightarrow \boxed{\text{a}} \longrightarrow \boxed{\text{b}} \longrightarrow \boxed{\text{c}} \longrightarrow$$

Solution:

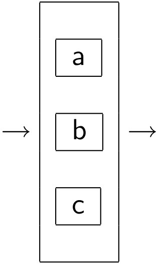
$$\begin{aligned} P(\text{System Works}) &= P(A \cap B \cap C) \\ &= P(A)\,P(B)\,P(C) \\ &= 0.95^3 = 0.857 \end{aligned}$$

PRACTICE

Problem 2: Suppose that

$$P(A) = P(B) = P(C) = 0.95.$$

Calculate the reliability of the system



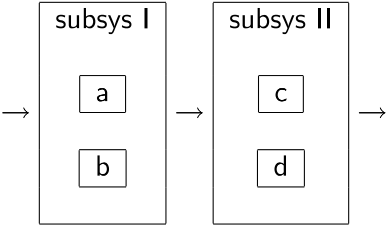
PROBLEM 2 (Solution)

$$
\begin{aligned}
P\left(\text{System works}\right) &= 1 - P\left(\text{System fails}\right) \\
&= 1 - P\left(A^c \cap B^c \cap C^c\right) \\
&= 1 - P\left(A^c\right) P\left(B^c\right) P\left(C^c\right) \\
&= 1 - \left(1 - P\left(A\right)\right)\left(1 - P\left(B\right)\right)\left(1 - P\left(C\right)\right) \\
&= 1 - 0.05^3 = 0.99988
\end{aligned}
$$

Problem 3: Suppose that

$$P(A) = P(B) = P(C) = P(D) = 0.95.$$

Calculate the reliability of the system



PROBLEM 3 (Solution)

$$
\begin{aligned}
P\left(\text{System works}\right) &= P\left(\text{subsys I works } \cap \text{ subsys II works}\right) \\
&= P\left(\text{subsys I works }\right) P\left(\text{subsys II works}\right) \\
&= \left[1 - P\left(\text{subsys I fails }\right)\right]\left[1 - P\left(\text{subsys II fails }\right)\right] \\
&= \left[1 - P\left(A^c \cap B^c\right)\right]\left[1 - P\left(C^c \cap D^c\right)\right] \\
&= \left[1 - P\left(A^c\right) P\left(B^c\right)\right]\left[1 - P\left(C^c\right) P\left(D^c\right)\right]
\end{aligned}
$$

$$= \left[1 - \left(1 - P\left(A\right)\right)\left(1 - P\left(B\right)\right)\right]\left[1 - \left(1 - P\left(C\right)\right)\left(1 - P\left(D\right)\right)\right]$$

$$= \left(1 - 0.05^2\right)^2 = 0.99501$$

CONDITIONAL INDEPENDENCE

Definition: We say that the events T_1, T_2, ..., T_n are conditionally independent given the event B if

$$P\left(T_{i_1} \cap T_{i_2} \cap \cdots \cap T_{i_k} \mid B\right) = P\left(T_{i_1} \mid B\right) P\left(T_{i_2} \mid B\right) \cdots P\left(T_{i_k} \mid B\right)$$

for all $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, and all $1 \leq k \leq n$.

For example, if $n = 3$, then

$$
\begin{aligned}
P(T_1 \cap T_2 \mid B) &= P(T_1 \mid B)\, P(T_2 \mid B) \\
P(T_1 \cap T_3 \mid B) &= P(T_1 \mid B)\, P(T_3 \mid B) \\
P(T_2 \cap T_3 \mid B) &= P(T_2 \mid B)\, P(T_3 \mid B) \\
P(T_1 \cap T_2 \cap T_3 \mid B) &= P(T_1 \mid B)\, P(T_2 \mid B)\, P(T_3 \mid B)
\end{aligned}
$$

Notes

- Conditional independence doesn't imply unconditional independence and vice versa
- Conditional independence given B doesn't imply conditional independence given B^c
- However, usually both conditional independences are assumed together in applications
- We will apply this concept in Bayesian probability updating

SEQUENTIAL BAYES' FORMULA (Bonus material)

Let S_i be the outcome of the ith test. For instance

$$S_1 = \left\{ \text{The } 1^{th} \text{ test is positive} \right\}$$

$$S_2 = \left\{ \text{The } 2^{th} \text{ test is negative} \right\}$$

$$S_3 = \left\{ \text{The } 3^{th} \text{ test is negative} \right\}$$

and so on

The outcomes S_i $(i = 1, 2, ..., n)$ are available in a sequential fashion.

Let $I_k = S_1 \cap S_2 \cap \cdots \cap S_k$ (data available at step k) and set

$$\pi_0 = P(D) \quad \text{Prior prob of an item being defective}$$

$$\pi_1 = P(D|I_1) = P(D|S_1) \quad \text{Posterior prob given } S_1$$

$$\pi_2 = P(E|I_2) = P(E|S_1 \cap S_2) \quad \text{Posterior prob given } S_1 \text{ and } S_2$$

$$\pi_3 = P(E|I_3) = P(E|S_1 \cap S_2 \cap S_3) \quad \text{Posterior prob given } S_1, S_2 \text{ and } S_3$$

and so on

CONDITIONAL INDEPENDENCE ASSUMPTION

Assume that the S_i $(i = 1, 2, ..., n)$ are independent given E and also given E^c.

Then, for $k = 1, 2, ..., n$

$$\pi_k = \frac{P(S_k|D)\,\pi_{k-1}}{P(S_k|D)\,\pi_{k-1} + P(S_k|D^c)\,(1 - \pi_{k-1})}$$

Proof

$$
\begin{aligned}
\pi_k &= \frac{P(I_k|D)\,\pi_0}{P(I_k|D)\,\pi_0 + P(I_k|D^c)\,(1-\pi_0)} \\
&= \frac{P(I_{k-1}\cap S_k|D)\,\pi_0}{P(I_{k-1}\cap S_k|D)\,\pi_0 + P(I_{k-1}\cap S_k|D^c)\,(1-\pi_0)} \\
&= \frac{P(S_k|D)\,P(I_{k-1}|D)\,\pi_0}{P(S_k|D)\,P(I_{k-1}|D)\,\pi_0 + P(S_k|D^c)\,P(I_{k-1}|D^c)\,(1-\pi_0)}
\end{aligned}
$$

(By the cond. independence assumption)

Proof (continued)

$$
\begin{aligned}
\pi_k &= \frac{P(S_k|D)\,P(I_{k-1}\cap D)}{P(S_k|E)\,P(I_{k-1}\cap E)+P(S_k|E^c)\,P(I_{k-1}\cap E^c)} \\
&= \frac{P(S_k|E)\,P(I_{k-1}\cap E)\,/\,P(I_{k-1})}{\left[P(S_k|E)\,P(I_{k-1}\cap E)+P(S_k|E^c)\,P(I_{k-1}\cap E^c)\right]/\,P(I_{k-1})} \\
&= \frac{P(S_k|E)\,\pi_{k-1}}{P(S_k|E)\,\pi_{k-1}+P(S_k|E^c)\,(1-\pi_{k-1})}, \qquad \pi_{k-1}=P(E|I_{k-1})
\end{aligned}
$$

Pseudo Code

- Input:
  - $(S_1, S_2, S_3, ..., S_n) = (1, 0, 1, ..., 0)$ (outcomes for the n tests)
  - $\pi = P(E)$ (prob of event of interest, for instance E= "the part is defective")
  - $p_k = P(S_k = +|E)$ $k = 1, 2, ..., n$ (Sensitivity of kth test)
  - $q_k = P(S_k = -|E^c)$ $k = 1, 2, ..., n$ (Specificity of kth test)

- Output $\pi_k = P(E|S_1 \cap S_2 \cap \cdots \cap S_k)$, $k = 1, 2, ..., n$

Example of Input:

$$n = 4, \quad \pi = 0.05$$

$$\text{Test Results} = (1, 1, 0, 1)$$

k	$p_k = P(1 \vert \text{Defective})$	$1 - q_k = P(1 \vert \text{Non Defective})$
1	$p_1 = 0.80$	$1 - q_1 = 0.05$
2	$p_2 = 0.78$	$1 - q_2 = 0.10$
3	$p_3 = 0.85$	$1 - q_3 = 0.20$
4	$p_4 = 0.82$	$1 - q_4 = 0.15$

Pseudo Code - Computation

Computation of π_k

1) Initialization: Set $\pi_0 = \pi$

2) k-step:

If $S_k = 1$, set $a = p_k$ and $b = 1 - q_k$

If $S_k = 0$, set $a = 1 - p_k$ and $b = q_k$

3) Computing π_k :

$$\pi_k = \frac{a\pi_{k-1}}{a\pi_{k-1} + b\left(1 - \pi_{k-1}\right)}$$

Example: Simple Spam Email Detection

- When you receive an email, your spam fillter uses Bayes rule to decide whether it is spam or not.

- Basic spam filters check whether some pre-specified words appear in the email; e.g.

  {diplomat,lottery,money,inheritance,president,sincerely,huge,...}.

- We consider n events W_i telling us whether the ith pre-specified word is in the message

- Let

$$E = \{ \text{ e-mail is spam } \}$$

$$W_i = \{ \text{ word i } \text{ is in the message } \}, \quad i = 1, 2, ..., n$$

- Assume that $W_1, W_2, ..., W_n$ are conditionally independent given E and also E^c.

- Human examination of a large number of messages is used estimate $\pi_0 = P(E)$
- The training data is also used to estimate $p_i = P(W_i|E)$ and $1 - q_i = P(W_i|E^c)$
- Let $I_n = S_1 \cap S_2 \cap \cdots \cap S_n$, where S_i is either W_i or W_i^c .
- The spam filter assumes that the W_i are conditionally independent (given E and given E^c) to compute

$$P(E|I_n) = \frac{P(I_n|E)\,P(E)}{P(I_n|E)\,P(E) + P(I_n|E^c)\,P(E^c)}$$

Sequential Updating

- The posterior probs $\pi_k = P(E|I_k) \ \ (k = 1, 2, ..., n-1)$ can be computed sequentially using the formula

$$\pi_k = P(E \mid I_k) = \frac{P(S_k|E)\,P(E|I_{k-1})}{P(S_k|E)\,P(E|I_{k-1}) + P(S_k|E^c)\,P(E^c|I_{k-1})}$$

$$= \frac{P(S_k|E)\,\pi_{k-1}}{P(S_k|E)\,\pi_{k-1} + P(S_k|E^c)\,(1-\pi_{k-1})}$$

- An early decision to classify the e-mail as spam can be made if $P(E \mid I_k)$ becomes too large (or too small).

Numerical Example

For a simple numerical example consider a case with

$$n = 8 \quad \text{words}, \quad P\,(\text{Spam}) = 0.10$$

and

Word	P(Word Present ∣ Spam)	P(Word Absent ∣ No Spam)
	Sensitivity	Specificity
W_1	0.74	0.98
W_2	0.83	0.88
W_3	0.88	0.89
W_4	0.75	0.99
W_5	0.82	0.85
W_6	0.73	0.89
W_7	0.77	0.93
W_8	0.86	0.92

Word	Word Status	$P(\text{Spam} \mid I_k)$
W_1	1	0.804
W_2	0	0.443
W_3	0	0.097
W_4	1	0.889
W_5	0	0.630
W_6	1	0.919
W_7	1	0.992
W_8	1	0.999