Let

\[ p(w_i|S) = \text{probability of word } w_i \text{ occurrence in spam e-mail.} \]
\[ p(w_i|\overline{S}) = \text{probability of word } w_i \text{ occurrence in non-spam e-mail.} \]

**I.** By Bayes Thm, the probability \( p(S|w_i) \) that an e-mail is spam if it contains \( w_i \) satisfies

\[
p(S|w_i) = \frac{p(w_i|S)p(S)}{p(w_i)} = \frac{p(w_i|S)p(S)}{p(w_i|S)p(S) + p(w_i|\overline{S})p(\overline{S})} = \frac{p(w_i|S)}{p(w_i|S) + \frac{1-p(S)}{p(S)}p(w_i|\overline{S})}\]

(For \( p(S) = p(\overline{S}) = \frac{1}{2} \), this reduces to \( p(S|w_i) = p(w_i|S)/(p(w_i|S) + p(w_i|\overline{S})) \).)

**II.** To combine the individual probabilities \( p_i \equiv p(S|w_i) \) for multiple pieces of evidence \( \vec{w} = \{w_1, \ldots, w_n\} \), again first use Bayes’ theorem in the form

\[
p(S|\vec{w}) = \frac{p(\vec{w}|S)}{p(\vec{w}|S) + \frac{1-p(S)}{p(S)}p(\vec{w}|\overline{S})}\]

This problem is underdetermined so assume statistical independence,*

\[
p(\vec{w}|S) = \prod_{i=1}^{n} p(w_i|S) \, , \quad p(\vec{w}|\overline{S}) = \prod_{i=1}^{n} p(w_i|\overline{S}) \, , \quad (I)\]

and again, using Bayes’ Thm, substitute \( p(w_i|S) = p_i p(w_i)/p(S) \) and \( p(w_i|\overline{S}) = (1-p_i)p(w_i)/p(\overline{S}) \) where \( p(\overline{S}|w_i) = 1 - p(S|w_i) = 1 - p_i \) to find

\[
p(S|\vec{w}) = \frac{\prod_i p_i \prod_i p(w_i)/p(S)}{\prod_i p_i \prod_i p(w_i)/p(S) + \frac{1-p(S)}{p(S)} \prod_i (1-p_i) \prod_i p(w_i)/1-p(S)} = \frac{\prod_i p_i}{\prod_i p_i + \left(\frac{p(S)}{1-p(S)}\right)^{n-1} \prod_i (1-p_i)} .\]

In the case \( p(S) = p(\overline{S}) = \frac{1}{2} \) (“flat prior”), this reduces to (“Bayes’ Rule”):

\[
p(S|\vec{w}) = \frac{\prod_i p_i}{\prod_i p_i + \prod_i (1-p_i)} \quad (B)\]

*Recall that independent events \( A \) and \( B \) satisfy \( p(A, B) = p(A)p(B) \), and this is equivalent both to \( p(A|B) = p(A) \) and to \( p(B|A) = p(B) \). Similarly, two events \( A \) and \( B \) are said
to be conditionally independent given a third event $C$, if $p(A, B|C) = p(A|C)p(B|C)$. This is equivalent both to $p(A|B, C) = p(A|C)$ and to $p(B|A, C) = p(B|C)$. The natural generalization to conditional independence of $n$ events is as in eqn. (I) above.

Another way to view this is to consider the “chain-rule”, following from multiple uses of the definition $p(A|B) = p(A, B)/p(B)$:

$$p(w_1, \ldots, w_n|S) = p(w_1, \ldots, w_n, S)/p(S) = p(w_2, \ldots, w_n|S, w_1)p(w_1|S)/p(S)$$
$$= p(w_1|S)p(w_2, \ldots, w_n|S, w_1)$$
$$= p(w_1|S)p(w_2|S, w_1)p(w_3, \ldots, w_n|S, w_1, w_2)$$
$$= p(w_1|S)p(w_2|S, w_1)p(w_3|S, w_1, w_2)p(w_4, \ldots, w_n|S, w_1, w_2, w_3)$$
$$= \ldots$$
$$= p(w_1|S)p(w_2|S, w_1)p(w_3|S, w_1, w_2)p(w_4|S, w_1, w_2, w_3)$$
$$\cdots p(w_n|S, w_1, w_2, w_3, \ldots, w_{n-1})$$

Assuming that every feature $w_i$ is conditionally independent of every other feature, i.e., $p(w_i|S, w_j) = p(w_i|S)$ ($i \neq j$), $p(w_i|S, w_j, w_k) = p(w_i|S)$ ($i \neq j, k$), \ldots, the above reduces to

$$p(w_1, \ldots, w_n|S) = p(w_1|S)p(w_2|S)p(w_3|S)p(w_4|S) \cdots p(w_n|S).$$