**The Binomial Naïve Bayes procedure**

The **Bernouilli Naïve Bayes procedure** in Section 8.2.2 looks at word occurrences; all that matters is whether a word is absent or present in a document. The **Binomial Naïve Bayes procedure** discussed here uses word frequencies. There it matters how often a word comes up in the document.

The prior probabilities are estimated the same way as explained previously, but the conditional joint probabilities  and  in equation (8.1) are estimated differently. For multinomial Naïve Bayes the marginal conditional probabilities of the occurrence of the *i*th word,  and , are obtained by dividing the **number of times** a given word occurs in documents of a given group (either group 1 or group 2) by the number of all word occurrences in that group; Laplace smoothing is implemented as well to avoid problems when a given word does not occur in one of the groups. Note this is different from the Bernouilli version which ignores how often a word is included in a document and works with occurrence indicators. Furthermore, the calculation of the posterior probability of the test case, where *n* is the total number of words of the test case and  are its number of occurrences of the *m* words, incorporates the word frequencies of the new case; the terms on the right-hand side of equation (8.2) are calculated from the multinomial distribution as



Note that the factorial terms in equation (8.1) cancel out so there is no need to calculate them.

An excellent write-up of this procedure and a detailed example are available on the Stanford natural language processing website, <https://nlp.stanford.edu/IR-book/html/htmledition/naive-bayes-text-classification-1.html>. The function **multinomial\_naive\_Bayes** of the R-library **naivebayes** can be used for the implementation, and an example of the R-code is shown on the website.