Attachment 1: Poisson binomial distribution

This section describes the Poisson binomial distribution, its relation to the simple binomial distribution, and a method for its rapid computation. In addition, the calculation formulas for the expected value and the variance for the random variables "number of correct answers" and "score" (in percent) are specified for both distributions.

First, in this section probabilities are not defined in percent, but, as usual in mathematics, in the interval $[0,1]$. They are therefore denoted by the letters π or Π instead of p. The interesting random variable $X \in \{0,1,2,\dots n\}$ is the number of correctly recognized words. If a subject has heard each of *n* words of a test list at the same level, this experiment yields a sequence of answers $(a) = (a_1, a_2, a_3, \dots, a_n)$ of *n* terms, that have the values $a_i = 1$ (word i recognized) or $a_i = 0$ (word i not recognized). Of interest is the probability $\Pi(X = k)$ that the participant recognized k words correctly. Assuming that for every word i there is a probability π_i that the word is recognized correctly, and that additionally the responses a_i are statistically independent, then Equation 13 defines the Poisson binomial distribution $B_\mathrm{g}(n, \vec{\pi}, k)$ with n words.

Equation 13:

$$\Pi(X = k) = B_\mathrm{g}(n, \vec{\pi}, k) = \sum_{\text{all } (a) \text{ with } \sum_{i=1}^{n} a_i \,=\, k} \; \prod_{i=1}^{n} \pi_i^{a_i} (1 - \pi_i)^{\bar{a}_i} \quad ,$$

$$\text{with} \quad \bar{a}_i = \begin{cases} 1 \text{ if } a_i = 0 \\ 0 \text{ if } a_i = 1 \end{cases} \quad \text{for } i = 0, 1, 2, \dots n. \qquad k = 0, 1, 2, \dots n.$$

The associated probabilities are summarized in the vector $\vec{\pi} = (\pi_1, \pi_2, \pi_3, \dots, \pi_n)^\mathrm{T} \in [0,\ 1]^n$. If all these probabilities are equal, i.e. $\pi_1 = \pi_2 = \cdots = \pi_n =: \pi_0$ holds, Equation 13 merges into the simple binomial distribution:

Equation 14

$$B(n, \pi_0, k) = \binom{n}{k} {\pi_0}^k (1 - \pi_0)^{n-k}$$

For general information on the mathematical properties of the Poisson binomial distribution, see [21].

The expected value and the variance of the random variable X binomially distributed according to Equation 14 are:

Equation 15

$$E_{n,\pi_0} = n\pi_0 \quad \text{and} \quad S^2_{n,\pi_0} = n\pi_0(1 - \pi_0)$$

For a random variable X distributed according to the Poisson binomial distribution Equation 13, the expected value and the variance are:

Equation 16

$$E_{n,\vec{\pi}} = \sum_{i=1}^{n} \pi_i \quad \text{and} \quad S^2_{n,\vec{\pi}} = \sum_{i=1}^{n} \pi_i (1 - \pi_i)$$

This follows directly from the fact that a random variable distributed according to Equation 13 is the sum of n independent Bernoulli-distributed random variables with the success probabilities π_i.

We consider the mean probability:

Equation 17

$$\bar{\pi} = \frac{1}{n}\sum_{i=1}^{n}\pi_i$$

The simple binomial distribution with $(n, \bar{\pi})$ and the Poisson binomial distribution with $(n, \vec{\pi})$ give the same expectation value:

Equation 18

$$E_{n,\vec{\pi}} \;=\; E_{n,\bar{\pi}} \;= n\bar{\pi}$$

The variances of the two distributions are generally not the same. Equations 15–17 yield:

Equation 19

$$S^2_{n,\vec{\pi}} \;=\; n\bar{\pi}(1-\bar{\pi}) - n\,s^2_{\vec{\pi}} \;=\; S^2_{n,\bar{\pi}} - n\,s^2_{\vec{\pi}} \quad \text{mit} \quad s^2_{\vec{\pi}} = \frac{1}{n}\sum_{i=1}^{n}(\pi_i - \bar{\pi})^2$$

The term $s^2_{\vec{\pi}}$ is the variance of π_i. If all π_i are equal, then $s^2_{\vec{\pi}} = 0$, and in this case the variance $S^2_{n,\vec{\pi}}$ of the random variable X is maximal. A variability of the individual probabilities thus results in a narrower distribution of the random variable. Figure 8 shows this for the example of a randomly chosen test list. The relative frequencies of word recognition measured for this test list were interpreted as probabilities π_i .

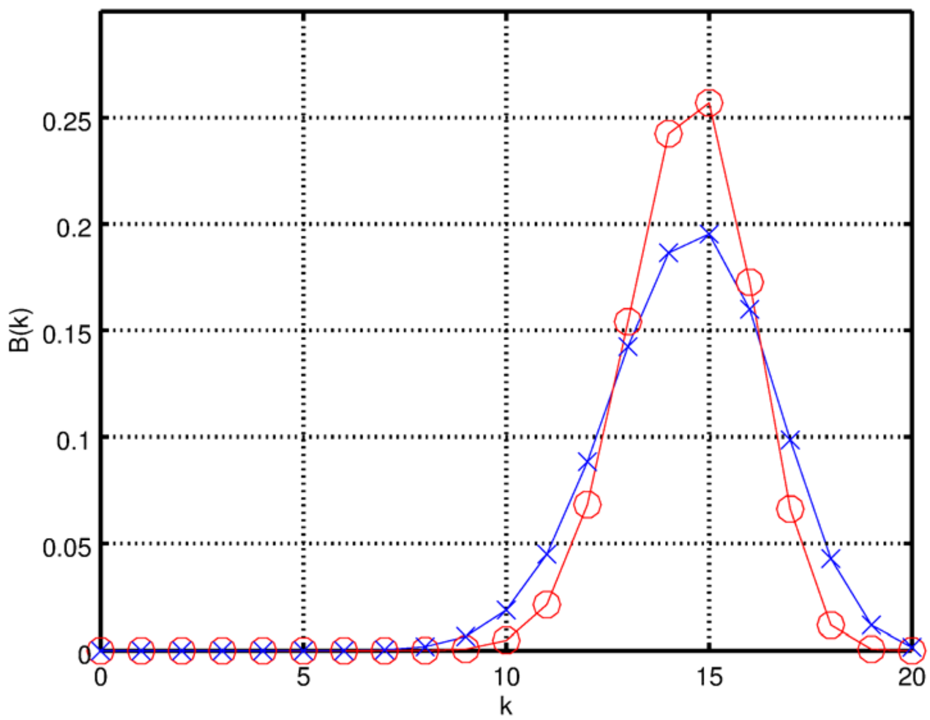

Figure 8: Binomial probability distribution $B(n, \bar{\pi}, k)$ (crosses) and Poisson binomial distribution $B_g(n, \vec{\pi}, k)$ (circles). Straight lines connect symbols for greater clarity. As a randomly-chosen example, the figure shows the measured data from list 2, level 3 (29.5 dB SPL), Group NH. Number of words $n = 20$. The observed relative frequencies of

Attachment 1 to: Holube I, Winkler A, Nolte-Holube R. Modeling the reliability of the Freiburg monosyllabic speech test in quiet with the Poisson binomial distribution. Does the Freiburg monosyllabic speech test contain 29 words per list? GMS Z Audiol (Audiol Acoust). 2020;2:Doc01. DOI: 10.3205/zaud000005

correct word recognition from this list and conditions are taken as the probabilities: $\vec{\pi} = (0.84, 1, 0.84, 0.79, 0, 0.95, 0.74, 0.79, 1, 0.89, 0.79, 0.63, 1, 0.21, 0.11, 1, 0.79, 1, 0.47, 0.63)^T$. *This results in* $\bar{\pi} = 0.72$, $E_{n,\vec{\pi}} = E_{n,\bar{\pi}} = 14.5, S^2_{n,\vec{\pi}} = 2.24$, $S^2_{n,\bar{\pi}} = 4.00$, $s^2_{\vec{\pi}} = 0.0878$. *Numbers have been rounded.*

The practical calculation of the Poisson binomial distribution can be done directly from Equation 13 by constructing a decision tree beginning with the first word $i = 1$, which at the stage i branches with the probability π_i to a correct result, and with the probability $1 - \pi_i$ to an incorrect result. However, the computation time for such a direct method increases very strongly with n, so that even for $n = 20$ the computation time for a complete distribution ($k = 0, 1, 2, \ldots n$) is of the order of minutes, depending on the PC used. Much faster is a method given by Hong [22], which is based on a discrete Fourier transformation (in the definition of ω, π occurs as circle number):

Equation 20

$$B_{\mathrm{g}}(n, \vec{\pi}, k) = \frac{1}{n+1} \sum_{l=0}^{n} \mathrm{e}^{\mathrm{i}\omega lk} \prod_{j=1}^{n} (1 - \pi_j + \pi_j \mathrm{e}^{\mathrm{i}\omega l}) \quad \text{with} \quad \omega = \frac{2\pi}{n+1}$$

In the present work, this method was used according to Equation 20 to calculate the Poisson binomial distribution.

In connection with speech recognition measurements, it is usually not the number of correctly understood words described by the random variable X that matters, but their share of the total number n. If this is expressed as a percentage, the new random variable $y = \frac{100}{n} X$ is obtained. This describes the score in percent. The probabilities π_i are also often given in percent. We therefore set $p_i = 100\pi_i$, $\bar{p} = 100\bar{\pi}$ etc.. The expectation value and the variance of the random variable y in the simple binomial distribution are thus:

Equation 21

$$E_{n,p_0}(y) = p_0 \quad \text{and} \quad \sigma^2_{n,p_0} = \frac{p_0(100 - p_0)}{n} = \left(\frac{100}{n}\right)^2 S^2_{n,\pi_0}$$

In a Poisson binomial distribution, expectation and variance of the random variable y are given by:

Equation 22

$$E_{n,\vec{p}}(y) = \frac{1}{n} \sum_{i=1}^{n} p_i = \bar{p} \quad \text{and} \quad \sigma^2_{n,\vec{p}} = \frac{1}{n^2} \sum_{i=1}^{n} p_i(100 - p_i) = \left(\frac{100}{n}\right)^2 S^2_{n,\vec{\pi}}$$

Equation 19 is thus for the score y as a random variable:

Equation 23

$$\sigma^2_{n,\vec{p}} = \sigma^2_{n,\bar{p}} - \frac{1}{n} s^2_{\vec{p}} \quad \text{with} \quad s^2_{\vec{p}} = \frac{1}{n} \sum_{i=1}^{n} (p_i - \bar{p})^2$$