INTRODUCTION

If one rejects a relatively well corroborated model, one immediately asks why the given data deviate from it. The answer can be given in different ways:

(I) The model is not adequate and must either be modified, generalized or rejected. If it has been derived deductively, then the error probably already lies in the approach, e.g., in the axioms that must be changed; if it has been reached inductively, then another model must be tried.

(II) The problem lies in the data itself. In connection with language data there are several possibilities:

(a) The text has been thoroughly corrected later by the author/editor so that the text is de facto a mixture of data.

(b) In texts there is always a need for originality that can be felt both in individual texts and in the whole genre. It just gives rise to genres. Frequently, this fact can be captured by simple variation of parameters of the same model, but need not immediately lead to the rejection of the model.

(c) If this need is too strong, then in particular cases the deviation can be too great and must be captured by a modification of the model.

(d) A language as a whole can – as is well known – develop and drift away from the ‘model’. Since language is always in a state of transition, one can expect stability of models of individual entities only for limited time intervals.

(e) The sample size is too small so that the trend cannot take a definite shape.

In individual cases, one does not know immediately which falsification is concerned. In general, one proceeds so that for each of four layers, namely (i) author; (ii) genre; (iii) language; and (iv) time, one assumes a uniform attractor. Here, an attractor is understood as being a space represented by a model – in the present study a distribution model. One assumes that each author has its own attractor, determined by the generating mechanism and boundary conditions (cf. e.g., Gaeta, 1994). In the same way, attractors developed for different genres in individual languages at different times. We can even assume that there are different generating mechanisms leading to quite unrelated models. Such a view automatically leads to the use of different (simple) models, even if one adheres to the existence of a unique law. In principle, one tries to start from a unique mechanism (cf. Wimmer et al., 1994; Wimmer & Altmann, 1994).
mann, 1996) and to use the same model for a set of homogeneous data. Unfortunately, it is not always possible to capture all data merely by variation of parameters. One always finds exceptions which are understood as signals of the author’s effort to leave the familiar attractor. If the deviations are small, the researchers – with or without the knowledge of causes which are hard to find out in texts – use the techniques of local or global modification of distributions which is found in all sciences and corresponds to the usual research process (cf. Lakatos, 1974): one maintains a theory until a new one replaces it and explains more than the original one.

Many researchers simultaneously use several models because they examine mixed data. But even in homogeneous data they sometimes present several models – as far as it is made possible by the existent software – because there is no special reason to choose a definite model inductively.

In this contribution we want to show the techniques of modification as applied to models of word length distribution. In this domain there are so many modifications that it is worth treating them in a unified way.

STATISTICAL APPROACH AND MODELLING

When measuring word length we have the following situation: in the given text of length $N$ (or in a sample of size $N$ from a dictionary), we measure the values $y_1, y_2, ..., y_N$, where $y_j$ is the length of the $j$th word ($j=1,2,...,N$) measured in terms of syllable, phoneme or letter numbers.

Let $X_i$ be the random variable ‘number of words of length $i’$ in the text or dictionary, $i=1,2,...,k$ (in some cases there is also $i=0$), then by simple summation of the values $y_1, y_2, ..., y_N$ we obtain the numbers $n_i$, designating the frequencies of words of length $i$ in text, $i=1,2,...,k$.

The vector $(n_1, n_2, ..., n_k)$ is the realisation of the random vector $(X_1, X_2, ..., X_k)$ in the given text (or dictionary), $k$ is the greatest word length, $N$ is the sample size (e.g., text length, sample size from the dictionary, etc.).

The common distribution of the vector $(X_1, X_2, ..., X_k)$ is considered to be multinomial with parameters $N$, $\pi_1, \pi_2, ..., \pi_k$, i.e.,

$$P(X_i = n_i, X_2 = n_2, ..., X_k = n_k) = \frac{N!}{n_1!n_2!...n_k!} \pi_1^{n_1}\pi_2^{n_2}...\pi_k^{n_k},$$

(1)

where $n_i \in \{0,1,...,N\}$, $i=1,2,...,k$ and $n_1 + n_2 + ... + n_k = N$.

The numbers $\pi_1, \pi_2, ..., \pi_k$ are the (theoretical) probabilities of the occurrences of words of length $i$ ($i=1,2,...,k$) and it holds that $0 < \pi_i < 1$ and $\pi_1 + \pi_2 + ... + \pi_k = 1$. However, in accordance with previous research we assume that $\{\pi_1, \pi_2, ..., \pi_k\}$ can be represented by different models that can be derived from appropriate approaches (cf. Wimmer et al., 1994; Wimmer & Altmann, 1996) under the assumption that the same model holds for ‘homogeneous’ data. The hypothesis that a certain model holds can be expressed mathematically as

$$H_0: \pi_i = P_i(\theta), \quad i=1,2,...,k, \quad \theta \in \mathbb{A} \subset \mathbb{R}^m,$$

where $\theta$ is the vector of parameters, $\mathbb{A}$ is the parameter space and $\mathbb{R}^m$ the $m$-dimensional Euclidian space.

An example of such a hypothesis is, e.g., Amentmann’s (1997) use of the positive negative binomial distribution in the form

$$H_0: \pi_i = \frac{\left(\theta_1 + i - 1\right)}{\theta_2} \theta_2^\theta (1 - \theta_2)^{i-1}, \quad i=1,2,...,k-1$$

$$\pi_k = \sum_{i=k}^\infty \frac{\left(\theta_1 + i - 1\right)}{\theta_2} \theta_2^\theta (1 - \theta_2)^{i-1},$$

$$\theta_1 > 0, \quad 0 < \theta_2 < 1,$$

where the probability of the greatest length class in the data, $k$, is determined as

$$\pi_k = 1 - \sum_{i=1}^{k-1} \pi_i.$$

A hypothesis of this kind can be tested. There is a number of test statistics with asymptotic properties which do not inspire much confidence on empirical researchers if they merely have samples with small sizes. Unfortunately, they
still have other disadvantages of a mathematical nature which could not be overcome as yet.

Let us consider in greater detail the test procedure starting from the assumption that if \( H_0 \) is valid the vector \( \theta_0 \in A \subset \mathbb{R}^m \), exists so that \( \pi_i = P_i(\theta_0), i = 1, 2, ..., k \). Under relatively general conditions about \( A \subset \mathbb{R}^m, \pi_i(\theta), \partial \pi_i(\theta)/\partial \theta_j \) (cf., e.g., Rao, 1973, 5e.2, 6a.2, 6b) one can show that for the so-called maximum likelihood (ML) estimator, \( \hat{\theta} = \hat{\theta}(X_1, X_2, ..., X_k) \), being the solution of the ML-equations

\[
\sum_{i=1}^{k} \frac{X_j \partial \pi_i(\theta)}{\partial \theta_j} \theta_j = 0, \quad j = 1, 2, ..., m
\]  

(2)

it holds that for \( N \to \infty \)

\[
\begin{pmatrix}
X_1 - N\pi_1(\hat{\theta}) \\
\sqrt{N\pi_1(\hat{\theta})}
\end{pmatrix}
\begin{pmatrix}
X_2 - N\pi_2(\hat{\theta}) \\
\sqrt{N\pi_2(\hat{\theta})}
\end{pmatrix}
\begin{pmatrix}
\vdots
\end{pmatrix}
\begin{pmatrix}
X_k - N\pi_k(\hat{\theta}) \\
\sqrt{N\pi_k(\hat{\theta})}
\end{pmatrix}
\to N(0_{k,1}; I - pp' - B(B'B)^{-1}B') \quad (3)
\]

i.e., the given vector converges in distribution to the normal distribution \( N(\mathbf{a}; \mathbf{C}) \), where \( \mathbf{a} \) is the mean and \( \mathbf{C} \) the covariance matrix. \( \mathbf{B} \) is a \( k \times m \)-matrix with rank \( \text{rank}(\mathbf{B}) = m \), whose \((i, j)\)-th element is

\[
\{B\}_{ij} = \frac{1}{\sqrt{\pi_i(\theta) \theta_j}} \partial \pi_i(\theta)/\partial \theta_j, \quad i = 1, 2, ..., k, \quad j = 1, 2, ..., m
\]

and, finally

\[
\hat{\theta} = \begin{pmatrix}
\sqrt{\pi_1(\hat{\theta})} \\
\sqrt{\pi_2(\hat{\theta})} \\
\vdots \\
\sqrt{\pi_k(\hat{\theta})}
\end{pmatrix}
\]

It also holds that

\[
\sqrt{N}
\begin{pmatrix}
\{\hat{\theta}_1 - \{\theta_0\}_1 \\
\{\hat{\theta}_2 - \{\theta_0\}_2 \\
\vdots \\
\{\hat{\theta}_m - \{\theta_0\}_m
\end{pmatrix}
\to N(0_{m,1}; (B'B)^{-1}), \quad (4)
\]

where \( \{\theta\}_j \) is the \( j \)th coordinate of the vector of parameters \( \theta \).

From the asymptotic behavior of the ML-estimators it follows (cf. Rao, 1973: 6a.2), that for sufficiently great \( N \) practically

\[
\begin{pmatrix}
X_1 - N\pi_1(\hat{\theta}) \\
\sqrt{N\pi_1(\hat{\theta})}
\end{pmatrix}
\begin{pmatrix}
X_2 - N\pi_2(\hat{\theta}) \\
\sqrt{N\pi_2(\hat{\theta})}
\end{pmatrix}
\begin{pmatrix}
\vdots
\end{pmatrix}
\begin{pmatrix}
X_k - N\pi_k(\hat{\theta}) \\
\sqrt{N\pi_k(\hat{\theta})}
\end{pmatrix}
\to N(0_{k,1}; I - pp' - \hat{\theta}(B'B)^{-1}B'), \quad (3a)
\]

and

\[
\hat{\rho} = \begin{pmatrix}
\sqrt{\pi_1(\hat{\theta})} \\
\sqrt{\pi_2(\hat{\theta})} \\
\vdots \\
\sqrt{\pi_k(\hat{\theta})}
\end{pmatrix}
\]

In the same way

\[
\begin{pmatrix}
\{\hat{\theta}_1 - \{\theta_0\}_1 \\
\{\hat{\theta}_2 - \{\theta_0\}_2 \\
\vdots \\
\{\hat{\theta}_m - \{\theta_0\}_m
\end{pmatrix}
\to N(0_{m,1}; (B'B)^{-1}). \quad (4a)
\]

For a normally distributed vector \( \xi \) with mean 0 and non-null idempotent covariance matrix \( \mathbf{V} \) it holds that \( \xi^2 \xi^2 \) is distributed as a \( \chi^2_{r(\mathbf{V})} \) [chi-square with \( r(\mathbf{V}) \) degrees of freedom, cf. Rao, 1973, 3b.4]. The matrix \( \mathbf{I} - pp' \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' \) is idempotent and non-null and of rank \( k-m \). Further, under the null hypothesis \( H_0 \) it holds according to (3) for \( N \to \infty \)

\[
\sum_{i=1}^{k} \frac{(X_i - N\pi_i(\hat{\theta}))^2}{N\pi_i(\hat{\theta})} \to \chi^2_{k-m-1}
\]

Thus, if for the realisation of \( (n_1, n_2, ..., n_k) \) and \( \hat{\theta}(n_1, n_2, ..., n_k) \) it holds that
\( \chi^2(\hat{\theta}) = \sum_{i=1}^{k} \frac{(n_i - N\pi_i(\hat{\theta}))^2}{N\pi_i(\hat{\theta})} > \chi^2_{a,m-1}(1 - \alpha) \) \hspace{1cm} (6)

\([\chi^2_{k-m-1}(1-\alpha)]\) is the \((1-\alpha)\)-quantile of the \(\chi^2_{k-m-1}\)-distribution, we reject \(H_0\) at the (asymptotical) level \(\alpha\).

Since we assume that in ‘homogeneous’ data the same model could be fitted, there is a possibility that the rejection of the model has been caused by a deformation of the data. Below we shall examine some of these deformations.

(A) Modification of the First Class

If we assume a deformation in the first class, we obtain the hypothesis (model)

\[ H_0^{(1)}: \pi_i = 1 - \gamma(1 - P_i(\theta)) \]

\[ \pi_i = \gamma P_i(\theta), \quad i = 2,3,...,k, \]

\[ \theta \in \Theta \subset \mathbb{R}^m, \quad \gamma \in (0, [1 - P_i(\theta)]^{-1}). \]

Here \(\gamma\) represents the deformation/modification parameter. For the modification of any class see chapter 4.

Estimation of the Parameters

In order to obtain the ML-equations (2) we proceed in practice as follows. We consider the so-called likelihood function

\[ L = \prod_{i=1}^{k} \pi_i(\gamma, \theta)^{X_i}, \]

which, in our case, yields

\[ L = \{1 - \gamma[1 - P_i(\theta)]\}^{\sum_{i=2}^{k} \gamma P_i(\theta)^{X_i}}. \]

Taking logarithms of \(L\), we obtain

\[ \ln L = X_i \ln[1 - \gamma(1 - P_i(\theta))] + \sum_{i=2}^{k} X_i \ln[\gamma P_i(\theta)]. \]

The maximum likelihood equations can be obtained by derivation of this formula according to individual parameters. For the parameter \(\gamma\) we obtain

\[ \frac{\partial \ln L}{\partial \gamma} = \frac{-X_i[1 - P_i(\hat{\theta})]}{1 - \hat{\gamma}[1 - P_i(\hat{\theta})]} + \sum_{i=2}^{k} \frac{X_i}{\hat{\gamma}} = 0 \] \hspace{1cm} (7a)

and for the other parameters

\[ \frac{\partial \ln L}{\partial \theta_j} = \frac{X_i \hat{\gamma}}{1 - \hat{\gamma}[1 - P_i(\hat{\theta})]} \frac{\partial P_i(\theta)}{\partial \theta_j}(\hat{\theta})|_{\hat{\theta}} = 0, \quad j = 1,2,...,m. \] \hspace{1cm} (7b)

\[ \hat{\gamma} = \frac{N - X_1}{N(1 - P_1(\hat{\theta}))} \] \hspace{1cm} (8a)

and after inserting (8a) in (7b)

\[ \frac{N - X_1}{1 - P_1(\hat{\theta})} \frac{\partial P_1(\hat{\theta})}{\partial \theta_j}(\hat{\theta})|_{\hat{\theta}} + \sum_{i=2}^{k} \frac{X_i}{P_i(\hat{\theta})} \frac{\partial P_i(\theta)}{\partial \theta_j}(\hat{\theta})|_{\hat{\theta}} = 0, \quad j = 1,2,...,m. \] \hspace{1cm} (8b)

We replace the variable \(X_i\) by the realised values \(n_i, (i = 1,2,...,k)\), i.e., by the observed frequencies and obtain the realisation (values) of the estimator \(\hat{\theta}\).

Test

The model can be tested in the usual way: if the value of the test criterion

\[ \chi^2(\hat{\gamma}, \hat{\theta}) = \frac{\{n_i - N\pi_i(\hat{\theta})\}^2}{N[1 - \hat{\gamma}(1 - P_i(\hat{\theta}))]} + \sum_{i=2}^{k} \frac{(n_i - N\hat{\gamma}P_i(\hat{\theta}))^2}{NP_i(\hat{\theta})} > \chi^2_{a,m-2}(1 - \alpha), \]

then we reject \(H_0^{(1)}\) at the (asymptotical) level \(\alpha\). In the opposite case, a model modified in the given way is not in contradiction with our measurements. In such cases one can compute the ‘degree of lack of fit’. To this end one uses, e.g., the measure of discrepancy (see Moore, 1984)

\[ d = \sum_{i=1}^{k} \frac{(\pi_i - Q_i)^2}{Q_i} \] \hspace{1cm} (9)

where \(\pi_i, (i = 1,2,...,k)\) are the true probabilities of words with length \(i\) and \(Q_i = \pi_i(\theta_0)\) \((i = 1,2,...,k)\) are the probabilities of words with length \(i\), when a particular model holds. The value of (9) can be estimated by means of
where $\theta^*(X_1, X_2, \ldots, X_k)$ is the minimum chi-square estimator. We obtain it as the solution of the minimization problem

$$
\theta^* = \arg \min_{\theta} \sum_{i=1}^{k} \frac{(X_i - N\pi_i(\theta))^2}{N\pi_i(\theta)}.
$$

It can be shown that for $N \to \infty$ the statistic

$$
\chi^2(\theta^*) = \sum_{i=1}^{k} \frac{(X_i - N\pi_i(\theta^*))^2}{N\pi_i(\theta^*)}
$$

is distributed as a chi-square with $k-m-1$ degrees of freedom ($m$ is the number of parameters, i.e., the dimension of the vector $\theta$). This means that if we use the ML-estimator in (6) or the minimum chi-square estimator in (12), we obtain the same asymptotical distribution for the test of fit.

It can also be shown (cf. Moore, 1994) that for $N \to \infty$ we obtain the convergence in distribution

$$
\sqrt{N} \left( \frac{\chi^2(\theta^*)}{N} - d \right) \to N(0; \tau^2)
$$

where

$$
\tau^2 = 4 \left\{ \frac{1}{2} \sum_{i=1}^{k} \frac{n_i^3}{Q_i} \right\} - \left\{ \sum_{i=1}^{k} \frac{n_i^2}{Q_i} \right\}^2
$$

which according to Moore (1984) holds only for the minimum chi-square estimator $\theta^*$. Since for a sufficiently great $N$

$$
\pi_i = \frac{n_i}{N}, \quad \pi(\theta) = Q_p
$$

we obtain the approximate value

$$
\tau^2 = 4 \left\{ \sum_{i=1}^{k} \frac{n_i^3}{N^2\pi_i(\theta)^3} - \left( \sum_{i=1}^{k} \frac{n_i^2}{N^2\pi_i(\theta)^2} \right)^2 \right\}
$$

and from (13)

$$
\frac{\chi^2(\theta^*)}{N} - d \approx N(0; \tau^2/N),
$$

i.e., also an approximate confidence interval for $d$.

Conclusions on the relation between the $d$ value and the linguistic considerations about the ‘distance of the model from reality’ can be obtained merely from empirical results. For example, a model is considered adequate if one does not reject the hypothesis about $d$ being zero. For other measures of discrepancy see Cressie and Read (1984).

(B) Modification of Two Classes

If the $H_0$ hypothesis is rejected according to (6), it can be hypothesized that the modification arose in two classes. If assumed that the deformation arose in the first two classes, then the hypothesis is

$$
H^{(2)}_0: \quad \pi_1 = P_1(\theta) + P_2(\theta) - \beta
$$

$$
\pi_2 = \beta
$$

$$
\pi_i = P_i(\theta), \quad i = 3, 4, \ldots, k;
$$

$$
\theta \in A \subset \mathbb{R}^m, \quad \beta \in (0, 1)
$$

The estimation of the parameters ($\beta^*, \theta^*$) can be performed, e.g., according to (11) (minimum chi-square) from

$$
(\beta^*, \theta^*) = \arg \min_{\beta \in (0, 1), \theta \in A} \left\{ \frac{(X_1 - N(P_1(\theta) + P_2(\theta) - \beta))^2}{N(P_1(\theta) + P_2(\theta) - \beta)} + \frac{(X_2 - N\beta)^2}{N\beta} - \sum_{i=3}^{k} \frac{(X_i - NP_i(\theta))^2}{NP_i(\theta)} \right\}
$$

setting the derivations according to the individual parameters equal to zero and testing the hypothesis $H_0^{(2)}$ according to (12). One can analogically obtain the measure of discrepancy (15).

(C) An Alternative Case

Another possibility of modification in case of rejection of $H_0$ in (6) is the following deformation

$$
H^{(0)}_0: \quad \pi_1 = 1 - \delta - \gamma(1 - P_1(\theta) - P_2(\theta))
$$

$$
\pi_2 = \delta
$$

$$
\pi_i = \gamma P_i(\theta), \quad i = 3, 4, \ldots, k;
$$

$$\theta \in A \subset \mathbb{R}^m, \quad \delta \in (0, 1),
$$

$$\gamma \in \left\{ 0, \frac{1 - \delta}{1 - P_1(\theta) - P_2(\theta)} \right\}.$$
i.e., we assume that the probabilities in classes \(i = 3,4,\ldots\) changed uniformly, while those in the first two classes changed ‘more substantially’. In this case the ML-estimation of \((\delta, \gamma, \theta)\) follows from the solution of equations

\[
\hat{\delta} = \frac{X_1 - X_2}{N}
\]

\[
\hat{\gamma} = \frac{N - X_1 - X_2}{N(1 - P_1(\theta) - P_2(\theta))}
\]

\[
\frac{\partial P_i(\theta)}{\partial \theta_j} \bigg|_{\theta = \theta_0} + \frac{\partial P_2(\theta)}{\partial \theta_j} \bigg|_{\theta = \theta_0} + \sum_{j=3}^m \frac{X_j}{P_j(\theta)} \frac{\partial P_i(\theta)}{\partial \theta_j} \bigg|_{\theta = \theta_0} = 0, \quad j = 1,2,\ldots,m
\]

In this case we can, for example, ask whether the deformation arose with \(\gamma = 1\), i.e., whether we have to do with the model \(H_0(2)\). If we know the ML-estimate of \((\delta, \hat{\gamma}, \hat{\theta})\), we can test, using (4a), whether \(\gamma = 1\) (see below). If we do not reject the assumption that \(\gamma = 1\), we can pass to the simpler model \(H_0(2)\).

**APPLICATIONS**

Let us illustrate the above procedures on examples from word length research.

**Example 1: The Singh Modification**

For a German scientific text Behrmann (1997) obtained the following values

<table>
<thead>
<tr>
<th>(i)</th>
<th>(n_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>454</td>
</tr>
<tr>
<td>2</td>
<td>293</td>
</tr>
<tr>
<td>3</td>
<td>154</td>
</tr>
<tr>
<td>4</td>
<td>87</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
</tr>
</tbody>
</table>

\(N = 1018\)

If one tries to fit to these data the positive Poisson distribution

\[
P_i(a) = \frac{a^i}{i!(e^a - 1)}, \quad i = 1,2,3,4,
\]

\[
P_5(a) = \sum_{i=5}^{\infty} \frac{a^i}{i!(e^a - 1)}.
\]

one obtains \(\chi^2_{3} = 14.41\), indicating that this model is not adequate. If one modifies this distribution according to hypothesis \(H_0(1)\), one obtains the positive Singh-Poisson distribution

\[
H_0(1): \quad \pi_1 = 1 - \gamma \left(1 - \frac{a}{e^a - 1}\right)
\]

\[
\pi_i = \gamma \frac{a^i}{i!(e^a - 1)}, \quad i = 2,3,4,
\]

\[
\pi_5 = \gamma \sum_{j=5}^{\infty} \frac{a^j}{j!(e^a - 1)}.
\]

If one uses for the estimation of the parameters the formulas (8a) and (8b), then one needs the derivations according to \(a\), namely

\[
\frac{\partial P_i(a)}{\partial a} = \frac{e^a - 1 - ae^a}{(e^a - 1)^2}, \quad i = 1,2,3,4,
\]

\[
\frac{\partial P_2(a)}{\partial a} = \frac{1}{(e^a - 1)^2} \left(1 + 1 + \frac{a^2}{2!} + \frac{a^3}{3!} + \frac{a^4}{4!}\right).
\]

From this one obtains the ML-equations (8a) and (8b) as follows

\[
\hat{\gamma} = \frac{1018 - 454 \left(1 - \frac{\hat{\gamma}}{1(e^\hat{\gamma} - 1)}\right)}{1018 - 454 \frac{e^{\hat{\gamma}} - 1}{1(e^{\hat{\gamma}} - 1)}}
\]

\[
= \frac{293}{3(e^{\hat{\gamma}} - 1)} \left(\frac{\hat{\gamma}(e^{\hat{\gamma}} - 1)^2}{2!} + \frac{\hat{\gamma}(e^{\hat{\gamma}} - 1)^2}{3!} + \frac{\hat{\gamma}(4e^{\hat{\gamma}} - 4 - e^{\hat{\gamma}})^2}{4!}\right)
\]

\[
+ \frac{30}{(e^{\hat{\gamma}} - 1)^2} \left(\frac{\hat{\gamma}^2(3e^{\hat{\gamma}} - 3 - 4e^{\hat{\gamma}})^2}{3!} + \frac{\hat{\gamma}^3(3e^{\hat{\gamma}} - 3 - 4e^{\hat{\gamma}})^2}{2!} + \hat{\gamma}^4 e^{\hat{\gamma}}\right) = 0.
\]

Equations (18a) and (18b) represent the ML-equations whose solution is

\[
\hat{\gamma} = 0.8822, \quad \hat{\alpha} = 1.7335,
\]

The second equation can be reordered and we obtain

\[
(1018 - 454) \left(\frac{\hat{\gamma}^2}{e^{\hat{\gamma}} - 1} + \frac{\hat{\gamma}^3}{3!(e^{\hat{\gamma}} - 1)}\right) + 293(2e^{\hat{\gamma}} - 2 - e^{\hat{\gamma}}) + 154(3e^{\hat{\gamma}} - 3 - 4e^{\hat{\gamma}})
\]

\[
+ 87(4e^{\hat{\gamma}} - 4 - e^{\hat{\gamma}}) + 30(\hat{\gamma}^2(3e^{\hat{\gamma}} - 3 - 4e^{\hat{\gamma}})^2) + \hat{\gamma}^3(3e^{\hat{\gamma}} - 3 - 4e^{\hat{\gamma}})^2 + \hat{\gamma}^4 e^{\hat{\gamma}} = 0.
\]

Equations (18a) and (18b) represent the ML-equations whose solution is
yielding according to \( H_0^{(1)} \) the values

\[
\begin{align*}
\pi_1 &= 0.4460 \quad N\pi_1 = 454.00 \\
\pi_2 &= 0.2844 \quad N\pi_2 = 289.53 \\
\pi_3 &= 0.1643 \quad N\pi_3 = 167.30 \\
\pi_4 &= 0.0712 \quad N\pi_4 = 72.50 \\
\pi_5 &= 0.0341 \quad N\pi_5 = 34.68.
\end{align*}
\]

The chi-square test criterion (6) yields

\[
\chi^2(\gamma, a) = 4.6286.
\]

Since the theoretical value \( \chi^2_2 (0.95) = 5.99 \) (> \( 4.6286 \)), we do not reject at the \( \alpha = 0.05 \) level the hypothesis that word length in the given text follows the positive Singh-Poisson distribution.

For the same test we also compute the estimate \((\gamma^*, a^*)\) using the minimum chi-square method (11). We have

\[
(\gamma^*, a^*) = \arg \min_{\gamma, a} \left\{ \frac{1}{N} \sum_{i=1}^{5} \left( \frac{n_i - N\gamma a_i}{N (e^a - 1)} - \frac{1}{e^a - 1} \right) \right\}^2 + \frac{1}{N} \sum_{i=1}^{5} \left( \frac{n_i - N\gamma \sum_{j=1}^{5} \frac{a_j}{j} (e^a - 1)}{N (e^a - 1)} \right)^2.
\]

Deriving the above equation according to \( \gamma \) and \( a \) and setting them equal to zero we obtain

\[
\gamma^* = 0.8808 \quad a^* = 1.7425,
\]

from which the values

\[
\begin{align*}
\pi_1 &= 0.4450 \quad N\pi_1 = 452.97 \\
\pi_2 &= 0.2838 \quad N\pi_2 = 288.91 \\
\pi_3 &= 0.1648 \quad N\pi_3 = 167.81 \\
\pi_4 &= 0.0718 \quad N\pi_4 = 73.10 \\
\pi_5 &= 0.0346 \quad N\pi_5 = 35.21
\end{align*}
\]

follow. The test criterion (12) yields

\[
\chi^2_2(\gamma^*, a^*) = 4.6102,
\]

which is in accordance with not rejecting (at the \( \alpha = 0.05 \) level) the hypothesis of adequacy of the positive Singh-Poisson distribution. Iterative fitting by means of FITTER (1994) yields \( \chi^2_2(\gamma^*, a^*) = 4.61 \) (Behrmann, 1997), i.e., merely a slight improvement of the fit.

Let us compute the discrepancy \( d \). It holds that

\[
\frac{\chi^2(\gamma^*, a^*)}{N} = \frac{4.6102}{1018} = 0.0045.
\]

According to (14) we have

\[
\tau^2 = 0.0192 \quad \text{and} \quad \frac{\tau^2}{N} = 0.0000189
\]

Thus according to (15) the approximate 95% confidence interval for \( d \) is

\[
\left( \frac{\chi^2_2(\gamma^*, a^*)}{N} - 1.96\sqrt{\frac{\tau^2}{N}}, \frac{\chi^2_2(\gamma^*, a^*)}{N} + 1.96\sqrt{\frac{\tau^2}{N}} \right),
\]

i.e., the interval \((-0.004; 0.013)\) contains the true value of \( d \) with (approximate) probability of 0.95. Since this interval contains zero, we cannot reject the assumption (at the \( \alpha = 0.05 \) level) that the discrepancy is zero. Thus we consider the discrepancy practically as zero.

**Example 2: An Alternative Modification**

For the distribution of word length in an Italian text, Hollberg (1997) obtained the following values

<table>
<thead>
<tr>
<th>( i )</th>
<th>( n_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>440</td>
</tr>
<tr>
<td>2</td>
<td>366</td>
</tr>
<tr>
<td>3</td>
<td>291</td>
</tr>
<tr>
<td>4</td>
<td>112</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

\( N = 1254 \)

After rejecting the hypothesis that these data follow the 1-displaced Poisson distribution

\[
H_0: \quad \pi_i = P_i(a) = \frac{e^{-a}a^{i-1}}{(i-1)!}, \quad i = 1, 2, \ldots, 6,
\]

\[
\pi_7 = P_7(a) = \sum_{j=7}^{n} \frac{e^{-a}a^{j-1}}{(j-1)!}
\]
we assume that the following modification is involved

\[ H_0(a) : \pi_i = 1 - \delta - \gamma(1 - P_i(a) - P_j(a)) \]

\[ \pi_2 = \delta \]

\[ \pi_i = \gamma P_i(a), \quad i = 3, 4, \ldots, 7 \]

\[ a > 0, \quad \delta \in (0, 1), \quad \gamma \in \left(0, \frac{1 - \delta}{1 - P_i(a) - P_j(a)}\right) \]

It holds that

\[ \frac{\partial P_i(a)}{\partial a} = -e^{-a}, \quad \frac{\partial P_i(a)}{\partial a} = e^{-a} a^{i-2} \left(1 - \frac{a}{i - 1}\right), \quad i = 2, 3, 4, 5, 6, \]

\[ \frac{\partial P_i(a)}{\partial a} = e^{-a} a^5 \frac{120}{i(i - 1)} . \]

The ML-equations (17) yield

\[ \dot{\delta} = \frac{366}{1254} \]

\[ \dot{\gamma} = \frac{1254 - 440 - 366}{1254(1 - e^{-\delta} - a e^{-a})} = 448 \frac{1254(1 - e^{-\delta} - \dot{a} e^{-a})}{1254(1 - e^{-\delta} - a e^{-a})} \]

\[ \frac{-448 e^{-\dot{a}}}{1 - e^{-\delta} - \dot{a} e^{-a}} + \sum_{i=3}^{6} \frac{n_i(i - 1)(1 - \hat{\dot{a}})}{\hat{\dot{a}}^i(i - 1)!} + \frac{n_i \hat{\dot{a}}^5}{120 \sum_{i=3}^{6} \hat{\dot{a}}^{i-1}(i - 1)!} = 0 \]

and their iterative solution gives

\[ \dot{\gamma} = 1.1042, \quad \dot{\delta} = 0.2919, \quad \hat{\dot{a}} = 1.1618. \]

Using these estimates we obtain the values

\[ \pi_1 = 0.3509, \quad N\pi_1 = 440.00 \]

\[ \pi_2 = 0.2919, \quad N\pi_2 = 366.00 \]

\[ \pi_3 = 0.2332, \quad N\pi_3 = 292.44 \]

\[ \pi_4 = 0.0903, \quad N\pi_4 = 113.25 \]

\[ \pi_5 = 0.0262, \quad N\pi_5 = 32.90 \]

\[ \pi_6 = 0.0061, \quad N\pi_6 = 7.64 \]

\[ \pi_7 = 0.0014, \quad N\pi_7 = 1.77 \]

Since

\[ \frac{\partial \pi_i(\gamma, \delta, a)}{\partial \gamma} = -1 + e^{-a} + \alpha e^{-a}, \quad \frac{\partial \pi_i(\gamma, \delta, a)}{\partial \delta} = 0, \]

\[ \frac{\partial \pi_i(\gamma, \delta, a)}{\partial \gamma} = e^{-a} a^{i-1} \left(1 - \frac{a}{i - 1}\right), \quad i = 3, 4, 5, 6, \]

\[ \frac{\partial \pi_i(\gamma, \delta, a)}{\partial \delta} = -1, \quad \frac{\partial \pi_i(\gamma, \delta, a)}{\partial \delta} = 1, \quad i = 3, 4, 5, 6,7, \]

\[ \frac{\partial \pi_i(\gamma, \delta, a)}{\partial a} = -\gamma a e^{-a}, \quad \frac{\partial \pi_i(\gamma, \delta, a)}{\partial a} = 0, \quad i = 3, 4, 5, 6, 7. \]

the matrix \( B \) looks like (19).

\[
\begin{pmatrix}
-1 & e^{-\dot{a}e^{-a}} & 1 & -\gamma a e^{-a} \\
\frac{1}{\sqrt{\pi_1}} & \frac{1}{\sqrt{\pi_1}} & \frac{1}{\sqrt{\pi_1}} & \frac{1}{\sqrt{\pi_1}} \\
0 & 0 & 0 & 0 \\
e^{-\dot{a}^2} & 0 & 0 & 0 \\
\frac{21}{\sqrt{\pi_3}} & \frac{21}{\sqrt{\pi_3}} & \frac{21}{\sqrt{\pi_3}} & \frac{21}{\sqrt{\pi_3}} \\
e^{-\dot{a}^3} & 0 & 0 & 0 \\
\frac{31}{\sqrt{\pi_5}} & \frac{31}{\sqrt{\pi_5}} & \frac{31}{\sqrt{\pi_5}} & \frac{31}{\sqrt{\pi_5}} \\
e^{-\dot{a}^4} & 0 & 0 & 0 \\
\frac{41}{\sqrt{\pi_7}} & \frac{41}{\sqrt{\pi_7}} & \frac{41}{\sqrt{\pi_7}} & \frac{41}{\sqrt{\pi_7}} \\
e^{-\dot{a}^5} & 0 & 0 & 0 \\
\frac{51}{\sqrt{\pi_9}} & \frac{51}{\sqrt{\pi_9}} & \frac{51}{\sqrt{\pi_9}} & \frac{51}{\sqrt{\pi_9}} \\
\frac{1 - \sum_{j=5}^{7} a^j/2}{\gamma} & 0 & 0 & \frac{\gamma^a e^{-a} \sqrt{\pi_7}}{2 \gamma} \end{pmatrix}
\]

By substituting the values of \( \dot{\gamma}, \dot{\delta}, \hat{\dot{a}} \) in (19) and using the values of \( \pi_i \) as computed above, we obtain the individual elements of the matrix \( \dot{B} \) according to (5) as follows:

\[ \{\dot{B}\}_{1,1} = \frac{-1 + e^{-1.618} + 1.1618 e^{-1.618}}{\sqrt{0.3509}} = -0.5461 \]

\[ \{\dot{B}\}_{2,1} = 0 \]

\[ \{\dot{B}\}_{3,1} = \frac{e^{-1.618} 1.1618^2}{21\sqrt{0.2323}} = 0.4373 \]

etc., so that finally we obtain the matrix \( \dot{B} \)

\[
\begin{pmatrix}
-0.5461 & -1.6881 & -0.6777 \\
0 & 1.8509 & 0 \\
0.4373 & 0 & 0.3484 \\
0.2722 & 0 & 0.4755 \\
0.1468 & 0 & 0.3959 \\
0.0707 & 0 & 0.2578 \\
0.0341 & 0 & 0.1629 \\
\end{pmatrix}
\]

This yields the matrix

\[
\frac{1}{N}(\dot{B})^{-1} \cdot \dot{B} = \frac{1}{1254} \begin{pmatrix} \dot{B} \end{pmatrix} = \begin{pmatrix}
12.7545 & -0.3223 & -8.5112 \\
-0.3223 & 0.2067 & 0.0000 \\
-8.5112 & 0.0000 & 6.8594 \end{pmatrix}
\]
According to (4a) it holds that
\[ \hat{\gamma} - \gamma \approx N \left( 0; \frac{12.7545}{1254} \right), \]
i.e., the (approximate) 95% confidence interval for \( \gamma \) is
\[ \left( \hat{\gamma} - 1.96 \sqrt{\frac{12.7545}{1254}}; \hat{\gamma} + 1.96 \sqrt{\frac{12.7545}{1254}} \right) \]

The interval
\[ (1.1042 - 0.1977; 1.1042 + 0.1977) = (0.907; 1.302) \]
contains the true value of \( \gamma \) with probability 0.95. Since this interval contains the value 1, one cannot reject (at the 0.05 level) the assumption (hypothesis) that \( \gamma = 1 \). Now, if one uses \( \gamma = 1 \), one obtains the hypothesis (model)

\[ \begin{align*}
\pi_1 &= P_1(a) + P_2(a) - \beta = e^{-a}(1 + a) - \beta, \\
\pi_2 &= \beta, \\
\pi_i &= P_i(a) = \frac{e^{-a}a^{i-1}}{(i-1)!}, \quad i = 3, 4, 5, 6, \\
\pi_7 &= P_7(a) = \sum_{i=7}^{\infty} \frac{e^{-a}a^{i-1}}{(i-1)!}, \\
a &> 0, \quad \beta \in (0, 1).
\end{align*} \]

The estimates \( (\beta^*, a^*) \) can be computed from (16) by means of the minimum chi-square method from the formula

\[ \begin{align*}
(\beta^*, a^*) &= \arg\min_{\beta, a} \left[ \frac{\left( n_1 - N(e^{-a} + ae^{-a} - \beta) \right)^2}{N(e^{-a} + ae^{-a} - \beta)} + \frac{(n_2 - N\beta)^2}{N\beta} + \\
&\quad + \frac{\sum_{i=3}^{3} \left( \frac{n_i - N\sum_{j=1}^{i-1} e^{-a}a^{j-1}}{(i-1)!} \right)^2}{N\sum_{j=1}^{3} e^{-a}a^{j-1} (j-1)!} \right] \\
\end{align*} \]

We obtain
\[ \beta^* = 0.2936, \quad a^* = 1.2446 \]
from which according to \( H_0^{(2)} \) the values
\[ \begin{align*}
\pi_1 &= 0.3530 \quad N\pi_1 = 442.62, \\
\pi_2 &= 0.2936 \quad N\pi_2 = 368.16 \\
\pi_3 &= 0.2231 \quad N\pi_3 = 279.78 \\
\pi_4 &= 0.0926 \quad N\pi_4 = 116.08 \\
\pi_5 &= 0.0288 \quad N\pi_5 = 36.12 \\
\pi_6 &= 0.0072 \quad N\pi_6 = 8.99 \\
\pi_7 &= 0.0018 \quad N\pi_7 = 2.26
\end{align*} \]
can be obtained. The value of (12) is
\[ \chi^2(\beta^*, a^*) = 5.0595 \]
and is smaller than \( \chi^2(0.95) = 9.49 \), which means that we do not reject the hypothesis \( H_0^{(2)} \) (at the \( \alpha = 0.05 \) level). This is in accordance with the statement that the given distribution of word lengths follows the 1-displaced Cohen-Poisson distribution (with a small transformation of the parameters). For the sake of completeness, we can indicate that according to (14)
\[ \tau^2 = 0.0078 \]
so that according to (15) we obtain the 95% confidence interval for \( d \) as
\[ (-0.001; 0.009). \]

Since this interval contains zero, one cannot reject at the \( \alpha = 0.05 \) level the hypothesis that \( d = 0 \) (as in the previous example).

**KINDS OF MODIFICATION**

Let us resume in what follows the kinds of modification used in linguistics and show their special cases.

Let \( N \) be the set of natural numbers, \( N_0 = N \cup \{0\} \) and \( T, U, W \) its defined subsets. Let us call
\[ \{P_x\}_{x \in W \subset N_0} \]
- the original distribution that will be modified locally or globally
\[ \{Q_x\}_{x \in U \subset N_0} \]
- the resulting (modified) distribution.

The basic types of modification are:
Type I (Modification of One Class)

\[ Q_x = \begin{cases} 
1 - \gamma \sum_{j \in T \setminus \{c\}} P_j \quad & \text{when } x = c \\
\gamma P_x \quad & \text{when } x \in T, x \neq c,
\end{cases} \quad c \in \mathbb{N}_0, \quad \gamma \in \left(0, \frac{1}{\sum_{j \in T} P_j}\right), \quad T \subset W, \quad U = T \cup \{c\}. \]

Examples:

1. **Positive Pandey-Poisson distribution** \((a,c,\alpha)\), where the original distribution is the **positive Poisson distribution** \((a)\):

\[ p_x = \frac{a^x}{x!(e^a - 1)}, \quad x \in W = \mathbb{N} \]

and the resulting distribution is (see Wimmer & Altmann, 1996):

\[ Q_x = \begin{cases} 
\frac{a^x}{x!(e^a - 1)}, \quad x = 1,2,\ldots, c-1, c+1, c+2, \ldots \\
1 - \alpha + \frac{a^x}{c!(e^a - 1)}, \quad x = c \\
\end{cases} \]

\[ \gamma = \alpha, \quad T = \mathbb{N}, \quad U = \mathbb{N} \]

2. **Positive Singh-Poisson distribution** \((a,\alpha)\) (see Gaeta, 1994; Best, 1996; Behrmann, 1997; Bartens & Zöbelin, 1997), also arising from the positive Poisson distribution \((a)\) and having the form

\[ Q_x = \begin{cases} 
1 - \alpha + \frac{a^x}{e^a - 1}, \quad x = 1 \\
\frac{a^x}{x!(e^a - 1)}, \quad x = 2,3,4,\ldots \\
\end{cases} \]

\[ \gamma = \alpha, \quad T = \mathbb{N}, \quad U = \mathbb{N}, \quad c = 1. \]

3. **Positive Cohen-Poisson distribution** \((a,\alpha)\) (see Laass, 1996; Zhu & Best, 1997), arising from the Poisson distribution \((a)\)

\[ p_x = e^{-a} \frac{a^x}{x!}, \quad x \in W = \mathbb{N}_0 \]

and having the form

\[ Q_x = \begin{cases} 
\frac{(1 - \alpha)a}{e^a - 1 - \alpha a} \quad & x = 1 \\
\frac{a^x}{x!(e^a - 1 - \alpha a)} \quad & x = 2,3,4,\ldots \\
\end{cases} \]

\[ \gamma = \frac{e^a}{e^a - 1 - \alpha a}, \quad T = \{2,3,4,\ldots\}, \quad U = \mathbb{N}, \quad c = 1. \]

4. **Extended positive binomial distribution** \((n, p, \alpha)\) (see Girzig, 1997; Uhlířová, 1997), arising from the binomial distribution \((n, p)\)

\[ p_x = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x \in W = \{0,1,2,\ldots,n\} \]

and having the form

\[ Q_x = \begin{cases} 
1 - \alpha, \quad x = 0 \\
\binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 2,3,\ldots,n \\
\end{cases} \]

\[ \gamma = \frac{\alpha}{1 - (1 - p)^n}, \quad T = \{1,2,3,\ldots,n\}, \quad U = 0,1,2,\ldots,n, \quad c = 0. \]

5. **Modified Conway-Maxwell-Poisson distribution 1** \((a, b, \alpha)\) (see Kim & Altmann, 1996), arising from the 1-displaced Conway-Maxwell-Poisson distribution \((a, b)\)

\[ p_x = \frac{a^{x-1}}{[(x-1)]^b K}, \quad x \in W = \mathbb{N} \]

\((K\text{ is the norming constant})\) and having the form

\[ Q_x = \begin{cases} 
\alpha \quad x = 1 \\
\frac{(1 - \alpha)a^{x-1}}{[(x-1)]^b K} \quad x = 2,3,4,\ldots \\
\end{cases} \]

\[ \gamma = 1 - \alpha, \quad T = \{2,3,\ldots\}, \quad U = \mathbb{N}, \quad c = 1. \]

6. **Modified Conway-Maxwell-Poisson distribution 2** \((a, b, \alpha)\) (see Kim & Altmann, 1996), arising from the zero-truncated (positive) Conway-Maxwell-Poisson distribution \((a, b)\)

\[ p_x = \frac{a^x}{(x!)^b Z}, \quad x \in W = \mathbb{N} \]

\((Z\text{ is the norming constant})\) and having the form
7. Positive Cohen-negative binomial distribution \((k, p, \alpha)\) (see Laass, 1996), arising from the negative binomial distribution \((k, p)\) and having the form

\[
P_x = \binom{k + x - 1}{x} p^k (1 - p)^x, \quad x \in W = \mathbb{N}_0
\]

and having the form

\[
Q_x = \begin{cases} 
\frac{(1 - \alpha)k(1 - p)p^k}{1 - p^k - ak(1 - p)p^k}, & x = 1 \\
\frac{k + x - 1}{x} p^q (1 - p)^x, & x = 2, 3, 4, \ldots
\end{cases}
\]

\[
\gamma = 1 - \alpha, \quad T = \{2, 3, \ldots\}, \quad U = \mathbb{N}, \quad c = 1.
\]

Type II (Modification of Two Classes)
The general form is

\[
Q_x = \begin{cases} 
1 - \delta - \gamma \sum_{j \in T \setminus \{c,d\}} P_{x_j} & \text{when } x = c, \\
\delta, & \text{when } x = d, \\
\gamma P_x & \text{when } x \in T, x \notin \{c,d\},
\end{cases}
\]

\[
\delta \in (0, 1), \quad c, d \in \mathbb{N}_0, \quad \gamma \in \left(0, \frac{1 - \delta}{\sum_{j \in T \setminus \{c,d\}} P_j}\right), \quad U = T \cup \{c,d\}, \quad T \subset W
\]

Examples:

1. Modified 1-displaced hyper-Poisson distribution \((a, b)\) (see Best & Medrano, 1997; Egbers et al., 1997), arising from the 1-displaced hyper-Poisson distribution \((a)\) and having the form

\[
P_x = \frac{a^{x-1}}{F_1(1; b; a)b^{(x-1)}}, \quad x \in W = \mathbb{N}
\]

and having the form

\[
Q_x = \begin{cases} 
P_x & x = 1, 2, 5, 6, \ldots \\
(1 - \alpha)P_x & x = 3 \\
P_3 + \alpha P_5 & x = 4,
\end{cases}
\]

\[
\gamma = 1, \quad T = \mathbb{N}, \quad U = \mathbb{N}, \quad c = 3, \quad d = 4.
\]

2. Modified zero-truncated Conway-Maxwell-Poisson distribution \((a, b)\) (see Kim & Altmann, 1996), arising from the above mentioned zero-truncated Conway-Maxwell-Poisson distribution \((a, b)\) and having the form of \(Q_x\) with \(\gamma = 1, \quad T = \mathbb{N}, \quad U = \mathbb{N}, \quad c = 3, \quad d = 4\).

3. 1-displaced Cohen-Poisson distribution \((a, \alpha)\) (see Hollberg, 1997), arising from the above mentioned 1-displaced Poisson distribution \((a)\) and having the form

\[
Q_x = \begin{cases} 
e^{-\alpha(1 + a\alpha)} & x = 1 \\
a e^{-\alpha(1 - \alpha)} & x = 2 \\
e^{-\alpha(x - 1)} & x = 3, 4, \ldots
\end{cases}
\]

\[
\gamma = 1, \quad \delta = a e^{-\alpha(1 - \alpha)}, \quad T = \{3, 4, \ldots\}, \quad U = \mathbb{N}, \quad c = 1, \quad d = 2.
\]

Type III (Modification of Three Classes)
The general form is

\[
Q_x = \begin{cases} 
1 - \beta - \gamma - \delta \sum_{j \in T \setminus \{c,d,e\}} P_{x_j} & \text{when } x = c, \\
\beta, & \text{when } x = d, \\
\gamma, & \text{when } x = e, \\
\delta P_x & \text{when } x \in T, x \notin \{c,d,e\},
\end{cases}
\]

\[
\beta, \gamma \in (0, 1), \quad c, d, e \in \mathbb{N}_0, \quad \beta + \gamma < 1, \quad \delta \in \left(0, \frac{1 - \beta - \gamma}{\sum_{j \in T \setminus \{c,d,e\}} P_j}\right),
\]

\[
U = T \cup \{c,d,e\}, \quad T \subset W
\]

Examples: in this way Uhlířová (1996) and Altmann, Erat and Hřebček (1996) introduced several modifications of the positive binomial distribution.

Type IV (Modification of 4 Classes)
The only case found in linguistics is the modification of the positive binomial distribution by Uhlířová (1997). The general form is

\[
Q_x = \begin{cases} 
1 - \beta - \gamma - \omega \sum_{j \in T \setminus \{c,d,e,f\}} P_{x_j} & \text{when } x = c, \\
\beta, & \text{when } x = d, \\
\gamma, & \text{when } x = e, \\
\delta, & \text{when } x = f, \\
\omega P_x & \text{when } x \in T, x \notin \{c,d,e,f\},
\end{cases}
\]
\[ \beta, \gamma, \delta \in (0,1), c,d,e,f \in \mathbb{N}_p, \beta + \gamma + \delta < 1, \omega \in \left( 0, \frac{1 - \beta - \gamma - \delta}{\sum_{j \in T \cap \{c,d,e,f\}} P_j} \right), \]

\[ U = T \cup \{c,d,e,f\}, T \subset W. \]

Models III and IV appeared only in a somewhat more ‘reduced’ form, i.e., with further restrictions for parameters or fewer parameters (cf. Altmann, Erat & Hřebíček, 1996).

The estimations and tests of models III and IV can be constructed analogically to those for Types I and II, as shown above. Using the reduced types III and IV one can also apply the Neyman-Pearson likelihood-ratio tests (cf. Rao, 1973, 6e.3).

The modified distributions also occur in mixed and further specialized forms (cf. Wimmer & Altmann, 1999).

Since the ML and the minimum chi-square estimates can frequently be computed merely iteratively, it can be recommended either to use simpler estimations or, if available, a ready-made software (e.g., Altmann-FITTER, 1994).

REFERENCES


