### Eine statistische Verteilung mit Vorbelegung. Anwendung auf mathematische Sprachanalyse.

Die mathematische Analyse der Bildung zusammengesetzter aus einfacheren Sprachelementen führt auf Probleme, bei denen drei Faktoren getrennt werden müssen: erstens Systembedingungen (z. B. gibt es keine nullsilbigen Wörter, keine Sätze mit mehr Wörtern als Silben, keine zwei konsekutiven Lücken zwischen Wörtern usf.), zweitens ein statistischer Untergrund, der allen Texten gemeinsam ist, drittens die rein autorspezifischen Eigenschaften der Formalstruktur eines Textes[1]).

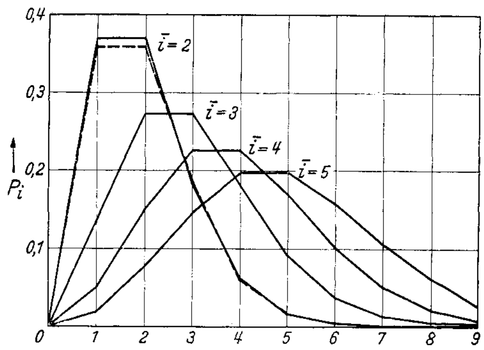


Fig. 1. Die $p_i'$-Verteilungen für die Mittelwerte $\bar{i}=2, 3, 4, 5$ bei einer Vorbelegung $b=1$. Für dieselbe Vorbelegung $b=1$ liefert das statistische Rechengerät für $\bar{i}=2$ die gestrichelte Kurve.

Wir suchen eine statistische Verteilung, die die beiden erstgenannten Faktoren umfaßt und daher die Abtrennung des dritten Faktors ermöglicht.

*Experimentell* erhält man solche Verteilungen mit Hilfe eines neuartigen statistischen Geräts[2]). Hat z. B. ein Text im Mittel zwei Silben je Wort ($\bar{i}=2$, wie bei gewissen lateinischen Texten), so belegt man zunächst die Zielzellen des Geräts mit einem Element (Kugel) je Zelle und läßt darauf aus jeder der Startzellen eine Kugel den statistischen Verteiler passieren. Die Mittelwerte der relativen Belegungshäufigkeiten der Zielzellen ergeben dann bei gehöriger Wiederholung die gesuchte

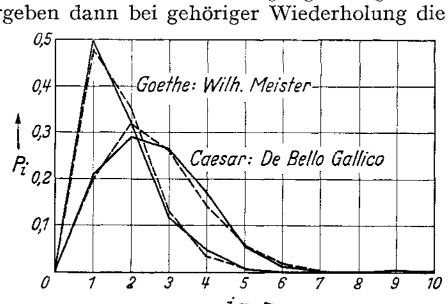


Fig. 2. Vergleich der Verteilungen für die relativen Häufigkeiten der Silben pro Wort für GOETHES Wilhelm Meister und CAESARS De Bello Gallico mit den auf dieselben Mittelwerte bezogenen theoretischen Verteilungen $p_i'$.

Verteilung. Beispielsweise ergab sich für $\bar{i}=2$ mit 37 Wiederholungen der gestrichelte Verlauf in Fig. 1.

*Theoretisch* ergibt sich für den Anteil der mit $i$ Kugeln belegten Zellen ($z_i'$) von den insgesamt $K$ Zielzellen bei einer Vorbelegung der Zielzellen mit je $b$ Kugeln offensichtlich

$$z_i' = K - \sum_{b}^{i-1} z_\nu' - \sum_{i-b+1}^{\infty} z_\nu\,, \qquad (1)$$

wenn $z_\nu/K$ die zugeordnete POISSON-Verteilung mit dem Mittelwert $\mu=\bar{i}-b$ bedeutet. Aus Gl. (1) errechnet sich die gesuchte Verteilung ($p_i'=z_i'/K$):

$$p_i' = \frac{e^{-(\bar{i}-b)}\cdot(\bar{i}-b)^{(i-b)}}{(i-b)!}\,; \qquad \sum_{i=0}^{\infty} p_i' = 1\,; \qquad (2)$$

mit den Momenten $\mu_1=\bar{i}$; $\mu_2=\bar{i}-b$; $\mu_3=\bar{i}-b$; $\mu_4=3(\bar{i}-b)^2+(\bar{i}-b)$; der Streuung $\sqrt{\bar{i}-b}$; der Schiefe $1/\sqrt{\bar{i}-b}$; der Flachheit $3+1/(\bar{i}-b)$.

Die Verteilung (2) ist in Fig. 1 für $b=1$ und $\bar{i}=2, 3, 4, 5$ dargestellt. In Fig. 2 ist die Verteilung der Silbenzahlen auf die Wörter von GOETHES Wilhelm Meister und CAESARS De Bello Gallico ausgezogen eingetragen. Die beiden gestrichelten Kurven sind nach Gl. (2) berechnet mit den Mittelwerten des entsprechenden Textes. Die Differenzen zwischen den theoretischen Kurven und den Kurven aus den Texten stellen die autorspezifischen Anteile dar.

Gl. (2) liefert die theoretische Deutung des Prozesses der Wortbildung aus Silben für 8 Sprachen[3]).

*Physikalisches Institut der Rheinisch-Westfälischen Technischen Hochschule Aachen.* W. FUCKS.

Eingegangen am 1. November 1954.

1) FUCKS, W.: Biometrika 39, 122 (1952). — Studium generale 6, 506 (1953). — Biometrika 41, 116 (1954).
2) FUCKS, W.: Naturwiss. 41, 57 (1954).
3) FUCKS, W.: Veröffentlichung der Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen (im Druck).

### Zur Abweichung von der BOLTZMANN-Verteilung.

Die Verteilung einer Teilchengesamtheit in einem Potentialfeld läßt sich bei vernachlässigbar kleiner Wechselwirkung durch den BOLTZMANNschen $e$-Satz beschreiben. Üben die Teilchen jedoch Kräfte aufeinander aus, so wird das Problem mathematisch sehr kompliziert und exakt nicht lösbar. In vorliegender Bemerkung sollen mit Hilfe der BORN-GREENschen Gleichungen unter Verwendung einfachster Superpositionsapproximationen Formeln für eine Mischung verschieden großer kugelförmiger Teilchen hergeleitet werden. Die Abweichung von der BOLTZMANN-Verteilung läßt sich in erster Näherung angeben.

Wir betrachten zwei Sorten kugelförmiger Teilchen, die wir durch die Indizes $a$ und $b$ unterscheiden. Ihre Ausschlußvolumina seien $v_{aa}$, $v_{ab}$ und $v_{bb}$. Dabei gilt $v_{aa}=(4\pi/3)\cdot a^3$; $v_{ab}=(\pi/6)\cdot(a+b)^3$; $v_{bb}=(4\pi/3)\,b^3$. Im vorliegenden Fall lauten die BORN-GREENschen Gleichungen[1])

$$\left.\begin{aligned}\operatorname{grad} n_a(x) = -\frac{1}{kT}\, n_a(x)\operatorname{grad}\Phi_a(x) - \frac{1}{kT}\int n_{aa}(x,y)\,\times\\ \times\operatorname{grad}\Phi_{aa}(x-y)\,dy - \frac{1}{kT}\int n_{ab}(x,y)\operatorname{grad}\Phi_{ab}(x-y)\,dy.\end{aligned}\right\}(1)$$

Eine entsprechende Gleichung gilt für $n_b(y)$. Es bedeuten in (1) $n_a(x)$ und $n_b(y)$ die gesuchten Konzentrationen, während $n_{aa}(x,y)$ und $n_{ab}(x,y)$ die radialen Verteilungsfunktionen sind. $x$ bzw. $y$ geben den Ort des Teilchens $a$ bzw. des Teilchens $b$ an. $dx$ bzw. $dy$ bedeuten Volumenelemente. Durch $\Phi_a(x)$ bzw. $\Phi_b(y)$ wird das Potentialfeld beschrieben. $\Phi_{aa}$ und $\Phi_{ab}$ geben die Wechselwirkungsenergien der Teilchen an. Jetzt machen wir für die radialen Verteilungsfunktionen die Ansätze

$$\left.\begin{aligned}n_{aa}(x,y) &= n_a(x)\,n_b(y)\,g_{aa}(x-y)\\ n_{ab}(x,y) &= n_a(x)\,n_b(y)\,g_{ab}(x-y)\,,\end{aligned}\right\}\qquad(2)$$

wobei $g_{aa}(x-y)$ gleich 0 oder 1 zu setzen ist, je nachdem $|x-y|$ kleiner oder größer als $a$ ist, während für $g_{ab}(x-y)$ der Sprung bei $(a+b)/2$ zu nehmen ist. Üben die Teilchen nur Kräfte bei Zusammenstößen aufeinander aus, so können wir für $\Phi_{aa}$ und $\Phi_{ab}$ die Potentiale starrer Kugeln einführen. Diese einfachen Näherungen gestatten es, die Gln. (1) zu integrieren. Für die Teilchensorte $a$ finden wir die Endformel, wenn $nv \ll 1$ ist,

$$n_a = n_a^0\, e^{-\Phi_a/kT}\,\frac{1+n_a^0 v_{aa}+n_b^0 v_{bb}+n_b^0(v_{bb}-v_{ab})\,e^{-\Phi_b/kT}}{1+n_a^0 v_{aa}\,e^{-\Phi_a/kT}+n_b^0 v_{bb}\,e^{-\Phi_b/kT}}\,. \qquad (3)$$

Die entsprechende Gl. (4) für die Teilchensorte $b$ erhält man aus (3) dadurch, daß man überall die Indizes $a$ und $b$ miteinander vertauscht ($v_{ba}=v_{ab}$).

$n_a^0$ und $n_b^0$ sind Konzentrationen an einem Nullniveau der Potentiale. Bei gleich großen Teilchen nimmt (3) die Form einer von EIGEN und WICKE aufgestellten Beziehung an, nur müssen wir an Stelle ihrer Besetzungszahl die Größe $1/v+n_a^0+n_b^0$ einführen. Formeln der Art (3) fanden Anwendung in der Theorie der starken Elektrolyte bei höheren Konzentrationen[2]). Die Berechtigung hierzu wird von uns näher untersucht.

*Rostock, Institut für Theoretische Physik.*
H. FALKENHAGEN und G. KELBG.

Eingegangen am 8. November 1954.

1) GREEN, H. S.: The Molecular Theory of Fluids, NHPC 1952.
2) FALKENHAGEN, H.: Elektrolyte, 2. Aufl. Leipzig: S. Hirzel 1953.