This article was downloaded by: [New York University]

On: 09 May 2015, At: 17:42

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Scandinavian Actuarial Journal

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/sact20>

Zur theorie der stabilität statistischer reihen

Al. A. Tschuprow

Published online: 22 Dec 2011.

To cite this article: Al. A. Tschuprow (1919) Zur theorie der stabilität statistischer reihen, Scandinavian Actuarial Journal, 1919:1, 80-133, DOI: [10.1080/03461238.1919.10414804](https://doi.org/10.1080/03461238.1919.10414804)

To link to this article: <http://dx.doi.org/10.1080/03461238.1919.10414804>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Zur Theorie der Stabilität statistischer Reihen.

Von

Al. A. Tschuprow.

DRITTE ABHANDLUNG.

Ueber den mittleren Fehler der »wesentlichen Schwankungscomponente«.

ERSTES KAPITEL.

I.

An N , beliebigen Verteilungsgesetzen folgenden, zufälligen Grössen — $x_1, x_2, \dots, x_N$ — werden N_i von einander unabhängige Versuche vorgenommen, — der erste Versuch an der Variablen x_1 , der zweite an x_2 u. s. w. Man setze:

$$E x_i = m_1^{(i)}; E x_i^2 = m_2^{(i)}; E x_i^k = m_k^{(i)},$$

$$E [x_i - m_1^{(i)}]^2 = m_2^{(i)} - [m_1^{(i)}]^2 = \mu_2^{(i)}; E [x_i - m_1^{(i)}]^k = \mu_k^{(i)},$$

$$\frac{1}{N} \sum_{i=1}^N m_1^{(i)} = m_{[1, N]}; \quad \frac{1}{N} \sum_{i=1}^N m_2^{(i)} = m_{[2, N]}; \quad \frac{1}{N} \sum_{i=1}^N m_k^{(i)} = m_{[k, N]},$$

$$\frac{1}{N} \sum_{i=1}^N \mu_2^{(i)} = \mu_{[2, N]}; \quad \frac{1}{N} \sum_{i=1}^N \mu_k^{(i)} = \mu_{[k, N]}.$$

Man bezeichne ferner mit x'_i den Wert, welchen die Variable x_i bei dem betreffenden Versuche erhält, und setze:

$$\frac{1}{N} \sum_{i=1}^N x'_i = x_{(N)},$$

$$E x_{(N)}^k = m_{k, (N)},$$

$$E [x_{(N)} - m_{1, (N)}]^k = \mu_{k, (N)}.$$

Da $m_{1, (N)} = E x_{(N)} = \frac{1}{N} E \sum_{i=1}^N x'_i = \frac{1}{N} \sum_{i=1}^N m_1^{(i)} = m_{[1, N]}$ ist, so lässt sich $\mu_{k, (N)}$ in der Form darstellen:

$$\begin{aligned} \mu_{k, (N)} &= E [x_{(N)} - m_{[1, N]}]^k = m_{k, (N)} + \\ &+ \sum_{h=1}^{k-2} (-1)^h C_k^h m_{[1, N]}^h m_{k-h, (N)} + (-1)^{k-1} (k-1) m_{[1, N]}^k. \end{aligned}$$

Andererseits hat man:

$$\begin{aligned} \mu_{[2, N]} &= \frac{1}{N} \sum_{i=1}^N [m_2^{(i)} - (m_1^{(i)})^2] = m_{[2, N]} - \frac{1}{N} \sum_{i=1}^N [m_1^{(i)}]^2 = \\ &= m_{[2, N]} - m_{[1, N]}^2 - \frac{1}{N} \sum_{i=1}^N [m_1^{(i)} - m_{[1, N]}]^2, \end{aligned}$$

$$\begin{aligned} \mu_{[k, N]} &= \frac{1}{N} \sum_{i=1}^N \sum_{h=0}^k (-1)^h C_k^h [m_1^{(i)}]^h m_{k-h}^{(i)} = \\ &= m_{[k, N]} + \sum_{h=1}^{k-2} (-1)^h C_k^h \frac{1}{N} \sum_{i=1}^N [m_1^{(i)}]^h m_{k-h}^{(i)} + \\ &+ (-1)^{k-1} (k-1) \frac{1}{N} \sum_{i=1}^N [m_1^{(i)}]^k. \end{aligned}$$

Da die Versuche gegenseitig unabhängig sind, so erhält man ferner:

82

$$m_{2,(N)} = Ex_{(N)}^2 = \frac{1}{N^2} \left\{ \sum m_2^{(i)} + 2 \sum_{i_1=1}^{N-1} \sum_{i_2=i_1+1}^N m_1^{(i_1)} m_1^{(i_2)} \right\} = \\ = m_{[1,N]}^2 + \frac{1}{N} \mu_{[2,N]} ,$$

$$m_{3,(N)} = Ex_{(N)}^3 = m_{[1,N]}^3 + \frac{3}{N} m_{[1,N]} \mu_{[2,N]} + \frac{1}{N^2} \mu_{[3,N]} ,$$

$$m_{4,(N)} = Ex_{(N)}^4 = m_{[1,N]}^4 + \frac{6}{N} m_{[1,N]}^2 \mu_{[2,N]} + \\ + \frac{1}{N^2} (4 m_{[1,N]} \mu_{[3,N]} + 3 \mu_{[2,N]}^2) + \frac{1}{N^3} \left\{ \frac{1}{N} \sum_{i=1}^N [\mu_i^{(i)} - 3(\mu_i^{(i)})^2] \right\} ,$$

$$\mu_{2,(N)} = E[x_{(N)} - m_{[1,N]}]^2 = \frac{1}{N} \mu_{[2,N]} ,$$

$$\mu_{3,(N)} = E[x_{(N)} - m_{[1,N]}]^3 = \frac{1}{N^2} \mu_{[3,N]} ,$$

$$\mu_{4,(N)} = E[x_{(N)} - m_{[1,N]}]^4 = \frac{3}{N^2} \mu_{[2,N]}^2 + \frac{1}{N^3} \left\{ \frac{1}{N} \sum_{i=1}^N [\mu_i^{(i)} - 3(\mu_i^{(i)})^2] \right\} .$$

II.

Setzt man

$$\frac{1}{N} \sum_{i=1}^N [x'_i - m_1^{(i)}]^k = \mu'_{[k,N]} ,$$

$$\frac{1}{N} \sum_{i=1}^N [x'_i - m_{[1,N]}]^k = \mu''_{[k,N]} ,$$

so ist

$$E \mu'_{[k,N]} = \mu_{[k,N]} ,$$

$$E \mu''_{[k,N]} = \frac{1}{N} \sum_{i=1}^N \sum_{h=0}^k C_{k,i}^h \mu_{k-h}^{(i)} [m_1^{(i)} - m_{[1,N]}]^h =$$

$$\begin{aligned}
&= \mu_{[k, N]} + k \frac{1}{N} \sum_{i=1}^N \mu_{k-1}^{(i)} [m_1^{(i)} - m_{[1, N]}] + \cdots + \\
&+ \frac{k(k-1)}{2} \frac{1}{N} \sum_{i=1}^N \mu_2^{(i)} [m_1^{(i)} - m_{[1, N]}]^{k-2} + \\
&\quad + \frac{1}{N} \sum_{i=1}^N [m_1^{(i)} - m_{[1, N]}]^k.
\end{aligned}$$

Die mathematische Erwartung von $\mu''_{[k, N]}$ lässt sich auch in einer anderen Form darstellen:

$$\begin{aligned}
E\mu''_{[k, N]} &= m_{[k, N]} + \\
&+ \sum_{h=1}^{k-2} (-1)^h C_k^h m_{[1, N]}^h m_{[k-h, N]} + (-1)^{k-1} (k-1) m_{[1, N]}^k.
\end{aligned}$$

Wenn alle Grössen $m_1^{(i)}$ untereinander gleich sind, so ist $\mu''_{[k, N]} = \mu'_{[k, N]}$, abgesehen davon, wie sich die Verteilungsgesetze im übrigen gestalten. Sind hingegen die Grössen $m_1^{(i)}$ nicht alle gleich, so ist $\mu''_{[k, N]} \neq \mu'_{[k, N]}$ und, im allgemeinen, auch $E\mu''_{[k, N]} \neq E\mu'_{[k, N]}$.

Bei $k=2$, hat man

$$E\mu''_{[2, N]} = m_{[2, N]} - m_{[1, N]}^2 = \mu_{[2, N]} + \frac{1}{N} \sum_{i=1}^N [m_1^{(i)} - m_{[1, N]}]^2,$$

woraus

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N [m_1^{(i)} - m_{[1, N]}]^2 &= E\{\mu''_{[2, N]} - \mu'_{[2, N]}\} = \\
&= E\left\{\frac{1}{N} \sum_{i=1}^N [x'_i - m_{[1, N]}]^2 - \frac{1}{N} \sum_{i=1}^N [x'_i - m_1^{(i)}]^2\right\}
\end{aligned}$$

folgt.

III.

Setzt man

$$\frac{1}{N} \sum_{i=1}^N [x'_i - x_{(N)}]^k = \nu'_{[k, N]},$$

$$E \nu'_{[k, N]} = \nu_{[k, N]},$$

so hat man

$$\begin{aligned} \nu'_{[1, N]} &= \nu'_{[1, N]} = 0, \\ (1) \quad \nu'_{[2, N]} &= E \frac{1}{N} \sum_{i=1}^N [x'_i - x_{(N)}]^2 = \\ &= \frac{1}{N} \sum_{i=1}^N [m_1^{(i)} - m_{[1, N]}]^2 + \frac{N-1}{N} \mu_{[2, N]}, \end{aligned}$$

$$\begin{aligned} (2) \quad \nu'_{[3, N]} &= E \frac{1}{N} \sum_{i=1}^N [x'_i - x_{(N)}]^3 = \frac{1}{N} \sum_{i=1}^N [m_1^{(i)} - m_{[1, N]}]^3 + \\ &+ \frac{(N-1)(N-2)}{N^2} \mu_{[3, N]} + \frac{3(N-2)}{N} \left[\frac{1}{N} \sum_{i=1}^N \mu_2^{(i)} m_1^{(i)} - \mu_{[2, N]} m_{[1, N]} \right]. \end{aligned}$$

Nach etwas langwierigeren Rechnungen findet man ferner:

$$\begin{aligned} (3) \quad E \nu'^2_{[2, N]} &= \left\{ \frac{1}{N} \sum_{i=1}^N [m_1^{(i)} - m_{[1, N]}]^2 \right\}^2 + \frac{(N-1)^2}{N^3} \mu_{[4, N]} + \\ &+ \frac{N^2 - 2N + 3}{N^4} \left\{ \left[\sum_{i=1}^N \mu_2^{(i)} \right]^2 - \sum_{i=1}^N [\mu_2^{(i)}]^2 \right\} + \\ &+ \frac{2(N-1)}{N} \mu_{[2, N]} \left\{ \frac{1}{N} \sum_{i=1}^N [m_1^{(i)} - m_{[1, N]}]^2 \right\} + \\ &+ \frac{4(N-1)}{N^2} \left\{ \frac{1}{N} \sum_{i=1}^N \mu_2^{(i)} m_1^{(i)} - \mu_{[2, N]} m_{[1, N]} \right\} + \frac{4}{N^2} \sum_{i=1}^N \mu_2^{(i)} [m_1^{(i)} - m_{[1, N]}]^2 \end{aligned}$$

und

$$\begin{aligned}
(4) \quad E[x'_{[2,N]} - x_{[2,N]}]^2 &= \\
&= E \left\{ \frac{1}{N} \sum_{i=1}^N [x'_i - x_{(N)}]^2 - \left[\frac{N-1}{N} \mu_{[2,N]} + \frac{1}{N} \sum_{i=1}^N [m_1^{(i)} - m_{[1,N]}]^2 \right] \right\}^2 = \\
&= \frac{(N-1)^2}{N^4} \sum_{i=1}^N [\mu_2^{(i)} - 3(\mu_2^{(i)})^2] + \frac{2(N-1)}{N^3} \sum_{i=1}^N [\mu_2^{(i)}]^2 - \\
&- \frac{2}{N^3} \sum_{i=1}^N [\mu_2^{(i)} - \mu_{[2,N]}]^2 + \frac{4}{N^2} \sum_{i=1}^N \mu_2^{(i)} [m_1^{(i)} - m_{[1,N]}]^2 + \\
&+ \frac{4(N-1)}{N^2} \left\{ \frac{1}{N} \sum_{i=1}^N \mu_2^{(i)} m_1^{(i)} - \mu_{[2,N]} m_{[1,N]} \right\}.
\end{aligned}$$

Im Falle, wenn alle Variablen einem und demselben Verteilungsgesetze folgen, erhält man aus (4):

$$(5) \quad E \left\{ \frac{1}{N-1} \sum_{i=1}^N [x'_i - x_{(N)}]^2 - \mu_2 \right\}^2 = \frac{2}{N-1} \mu_2^2 + \frac{1}{N} [\mu_2 - 3\mu_2^2].$$

ZWEITES KAPITEL.

I.

1) Wenn die N Variablen in r Serien von je n , einem und demselben Verteilungsgesetze folgenden, Variablen zerfallen, erhält man aus (1):

$$(6) \quad E \frac{1}{nr} \sum_{i=1}^{nr} [x'_i - x_{(nr)}]^2 = \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 + \frac{nr-1}{nr} \mu_{[2,r]},$$

wobei durch $m_{[1,r]}$ und $\mu_{[2,r]}$ die arithmetischen Durchschnitte der die r Verteilungsgesetze charakterisierenden Größen $m_1^{(i)}$ und $\mu_2^{(i)}$ bezeichnet werden, so dass $m_{[1,r]} = \frac{1}{r} \sum_{i=1}^r m_1^{(i)}$,

$$\mu_{[2,r]} = \frac{1}{r} \sum_{i=1}^r \mu_2^{(i)}.$$

Bezeichnet man ferner mit $x'_{i,j}$ den Werth, welchen beim j -ten Versuche der i -ten Serie die betreffende Variable annimmt, und setzt $z_i = \frac{1}{n} \sum_{j=1}^n x'_{i,j}$, so hat man:

$$\frac{1}{r} \sum_{i=1}^r z_i = \frac{1}{nr} \sum_{i=1}^r \sum_{j=1}^n x'_{i,j} = x_{(nr)},$$

$$E z_i = E x_i = m_1^{(i)}, \quad \frac{1}{r} \sum_{i=1}^r E z_i = m_{[1,r]},$$

$$E z_i^2 = \frac{1}{n} E x_i^2 + \frac{n-1}{n} [E x_i]^2 = \frac{1}{n} \mu_2^{(i)} + [m_1^{(i)}]^2,$$

$$E z_i^2 - [E z_i]^2 = \frac{1}{n} \mu_2^{(i)},$$

$$\frac{1}{r} \sum_{i=1}^r [E z_i^2 - (E z_i)^2] = \frac{1}{nr} \sum_{i=1}^r \mu_2^{(i)} = \frac{1}{n} \mu_{[2,r]},$$

$$(7) \quad E \frac{1}{r} \sum_{i=1}^r [z_i - x_{(nr)}]^2 = \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 + \frac{r-1}{nr} \mu_{[2,r]}.$$

Aus (6) und (7) erhält man:

$$(8) \quad \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 = \\ = E \left\{ \frac{nr-1}{(n-1)r} \frac{1}{r} \sum_{i=1}^r [z_i - x_{(nr)}]^2 - \frac{r-1}{r(n-1)} \frac{1}{nr} \sum_{i=1}^{nr} [x_i - x_{(nr)}]^2 \right\}.$$

Aus

$$E \frac{1}{n-1} \sum_{j=1}^n [x'_{i,j} - z_i]^2 = \mu_2^{(i)},$$

findet man anderseits:

$$(9) \quad \mu_{[2,r]} = E \frac{1}{r(n-1)} \sum_{i=1}^r \sum_{j=1}^n [x'_{i,j} - z_i]^2$$

und durch Substitution in (7):

$$(10) \quad \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 = \\ = E \left\{ \frac{1}{r} \sum_{i=1}^r [z_i - x_{(nr)}]^2 - \frac{r-1}{r(n-1)} \frac{1}{nr} \sum_{i=1}^r \sum_{j=1}^n [x'_{i,j} - z_i]^2 \right\}.$$

Dies ist jedoch kein neuer selbständiger Weg zur empirischen Bestimmung des Wertes von $\frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2$, da nicht nur die mathematischen Erwartungen der in (8) und in (10) zur rechten Seite stehenden Ausdrücke gleich, sondern auch die Ausdrücke selbst identisch gleich sind, wie man sich leicht überzeugen kann.

2) Setzt man

$$\frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 = M, \\ \frac{1}{r} \sum_{i=1}^r [z_i - x_{(nr)}]^2 = M', \\ \frac{r-1}{r(n-1)} \frac{1}{nr} \sum_{i=1}^r \sum_{j=1}^n [x'_{i,j} - z_i]^2 = M'',$$

so ist

$$E M' = M + \frac{r-1}{nr} \mu_{[2,r]},$$

$$E M'' = \frac{r-1}{nr} \mu_{[2,r]},$$

$$M = E \{M' - M''\}.$$

Man findet ferner leicht

$$\begin{aligned}
(10') \quad E[M' - E M']^2 &= \\
&= \frac{(r-1)^2}{n^3 r^3} \frac{1}{r} \sum_{i=1}^r [\mu_1^{(i)} - 3(\mu_2^{(i)})^2] + \frac{2(r-1)}{n^2 r^2} \frac{1}{r} \sum_{i=1}^r [\mu_2^{(i)}]^2 - \\
&- \frac{2}{n^2 r^2} \frac{1}{r} \sum_{i=1}^r [\mu_2^{(i)} - \mu_{[2, r]}]^2 + \frac{4}{n r} \frac{1}{r} \sum_{i=1}^r \mu_2^{(i)} [m_1^{(i)} - m_{[1, r]}]^2 + \\
&+ \frac{4(r-1)}{n^2 r^2} \left[\frac{1}{r} \sum_{i=1}^r \mu_2^{(i)} m_1^{(i)} - \mu_{[3, r]} m_{[1, r]} \right], \\
E[M'' - E M'']^2 &= \\
&= \frac{(r-1)^2}{n^3 r^3} \frac{1}{r} \sum_{i=1}^r [\mu_1^{(i)} - 3(\mu_2^{(i)})^2] + \frac{2(r-1)^2}{r^3 n^2 (n-1)} \frac{1}{r} \sum_{i=1}^r [\mu_2^{(i)}]^2
\end{aligned}$$

und nach etwas beschwerlicheren Rechnungen:

$$\begin{aligned}
E[M' - E M'] \cdot [M'' - E M''] &= \frac{(r-1)^2}{n^3 r^3} \frac{1}{r} \sum_{i=1}^r [\mu_1^{(i)} - 3(\mu_2^{(i)})^2] + \\
&+ \frac{2(r-1)}{n^2 r^2} \left[\frac{1}{r} \sum_{i=1}^r \mu_2^{(i)} m_1^{(i)} - \mu_{[3, r]} m_{[1, r]} \right].
\end{aligned}$$

Da nun aber

$$\begin{aligned}
E\{[M' - M''] - M\}^2 &= E[M' - E M']^2 + \\
&+ E[M'' - E M'']^2 - 2E\{[M' - E M'] \cdot [M'' - E M'']\},
\end{aligned}$$

so erhält man schliesslich:

$$\begin{aligned}
(11) \quad E\{[M' - M''] - M\}^2 &= \\
&= \frac{2(r-1)(nr-1)}{r^3 n^2 (n-1)} \frac{1}{r} \sum_{i=1}^r [\mu_2^{(i)}]^2 - \frac{2}{n^2 r^2} \frac{1}{r} \sum_{i=1}^r [\mu_2^{(i)} - \mu_{[2, r]}]^2 + \\
&+ \frac{4}{n r} \frac{1}{r} \sum_{i=1}^r \mu_2^{(i)} [m_1^{(i)} - m_{[1, r]}]^2.
\end{aligned}$$

II.

1) Man nehme nun an, dass an der Variablen $x_1 - n_1$ Versuche vorgenommen werden, an der Variablen $x_2 - n_2$ Versuche u. s. w., und setze $n_1 + n_2 + \dots + n_r = s$.

Betrachtet man alle s Versuche, als ein ganzes, so hat man:

$$m_{[1, s]} = \frac{1}{s} \sum_{i=1}^r n_i m_1^{(i)},$$

$$m_{[2, s]} = \frac{1}{s} \sum_{i=1}^r n_i m_2^{(i)},$$

$$\mu_{[2, s]} = \frac{1}{s} \sum_{i=1}^r n_i \mu_2^{(i)},$$

$$E x_{(s)} = m_{[1, s]},$$

$$E x_{(s)}^2 = m_{[1, s]}^2 + \frac{1}{s} \mu_{[2, s]},$$

$$(12) \quad E \frac{1}{s} \sum_{i=1}^s [x'_i - x_{(s)}]^2 = \frac{1}{s} \sum_{i=1}^r n_i [m_1^{(i)} - m_{[1, s]}]^2 + \frac{s-1}{s} \mu_{[2, s]}.$$

Setzt man anderseits

$$z_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x'_{i,j},$$

$$z_{(r)} = \frac{1}{r} \sum_{i=1}^r z_i,$$

so erhält man:

$$\frac{1}{s} \sum_{i=1}^r n_i z_i = x_{(s)},$$

$$E z_i = m_1^{(i)}; \quad \frac{1}{r} \sum_{i=1}^r E z_i = m_{[1,r]}; \quad \frac{1}{s} \sum_{i=1}^r n_i m_1^{(i)} = m_{[1,s]},$$

$$E z_i^2 = [m_1^{(i)}]^2 + \frac{1}{n_i} \mu_2^{(i)}; \quad E z_i^2 - [E z_i]^2 = \frac{1}{n_i} \mu_2^{(i)}$$

und, bei $n_i \geq 2$,

$$E \frac{1}{n_i - 1} \sum_{j=1}^{n_i} [x'_{i,j} - z_i]^2 = \mu_2^{(i)},$$

$$(13) \quad E \frac{1}{s} \sum_{i=1}^r \frac{n_i}{n_i - 1} \sum_{j=1}^{n_i} [x'_{i,j} - z_i]^2 = \frac{1}{s} \sum_{i=1}^r n_i \mu_2^{(i)} = \mu_{[2,s]},$$

$$E \frac{1}{r} \sum_{i=1}^r \frac{1}{n_i - 1} \sum_{j=1}^{n_i} [x'_{i,j} - z_i]^2 = \frac{1}{r} \sum_{i=1}^r \mu_2^{(i)} = \mu_{[2,r]},$$

$$E \frac{1}{r} \sum_{i=1}^r \frac{1}{n_i(n_i - 1)} \sum_{j=1}^{n_i} [x'_{i,j} - z_i]^2 = \frac{1}{r} \sum_{i=1}^r \frac{1}{n_i} \mu_2^{(i)}.$$

Man findet ferner mühelos:

$$(14) \quad E \sum_{i=1}^r [z_i - z_{(r)}]^2 = \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 + \frac{r-1}{r} \sum_{i=1}^r \frac{1}{n_i} \mu_2^{(i)},$$

$$(15) \quad E \sum_{i=1}^r n_i [z_i - x_{(s)}]^2 = \sum_{i=1}^r n_i [m_1^{(i)} - m_{[1,s]}]^2 + \sum_{i=1}^r \mu_2^{(i)} - \frac{1}{s} \sum_{i=1}^r n_i \mu_2^{(i)},$$

$$(16) \quad E \sum_{i=1}^r [z_i - x_{(s)}]^2 = \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 + \sum_{i=1}^r \frac{1}{n_i} \mu_2^{(i)} - \frac{2}{s} \sum_{i=1}^r \mu_2^{(i)} +$$

$$+ \frac{r}{s^2} \sum_{i=1}^r n_i \mu_2^{(i)} + 2r m_{[1,s]} [m_{[1,s]} - m_{[1,r]}],$$

$$(17) \quad E \sum_{i=1}^r n_i [z_i - z_{(r)}]^2 = \sum_{i=1}^r n_i [m_1^{(i)} - m_{[1,r]}]^2 + \frac{r-2}{r} \sum_{i=1}^r \mu_2^{(i)} +$$

$$+ \frac{s}{r^2} \sum_{i=1}^r \frac{1}{n_i} \mu_2^{(i)}.$$

Aus (14) und (13) erhält man:

$$(18) \quad \frac{1}{r} \sum_{i=1}^r [m_i^{(i)} - m_{[1,r]}]^2 = \\ = E \left\{ \frac{1}{r} \sum_{i=1}^r [z_i - z_{(r)}]^2 - \frac{r-1}{r^2} \sum_{i=1}^r \frac{1}{n_i(n_i-1)} \sum_{j=1}^{n_i} [x'_{i,j} - z_i]^2 \right\},$$

und aus (12) und (13)

$$(19) \quad \frac{1}{s} \sum_{i=1}^r n_i [m_i^{(i)} - m_{[1,s]}]^2 = \\ = E \left\{ \frac{1}{s} \sum_{i=1}^s [x'_i - x_{(s)}]^2 - \frac{s-1}{s^2} \sum_{i=1}^r \frac{n_i}{n_i-1} \sum_{j=1}^{n_i} [x'_{i,j} - z_i]^2 \right\},$$

oder auch aus (15) und (13)

$$(20) \quad \frac{1}{s} \sum_{i=1}^r n_i [m_i^{(i)} - m_{[1,s]}]^2 = \\ = E \left\{ \frac{1}{s} \sum_{i=1}^r n_i [z_i - x_{(s)}]^2 - \frac{1}{s^2} \sum_{i=1}^r \frac{s-n_i}{n_i-1} \sum_{j=1}^{n_i} [x'_{i,j} - z_i]^2 \right\}.$$

Die Formel (18) zeigt, dass der mittlere Fehler der mathematischen Erwartungen der Grössen z_i stets kleiner ist, als die mathematische Erwartung des mittleren Fehlers der Grössen z_i . Was die Formeln (19) und (20) anbelangt, so lässt es sich leicht nachweisen, dass nicht nur die mathematischen Erwartungen der auf der rechten Seite stehenden Ausdrücke gleich, sondern auch die Ausdrücke selber identisch gleich sind: die beiden Wege zur empirischen Bestimmung von $\frac{1}{s} \sum_{i=1}^r n_i [m_i^{(i)} - m_{[1,s]}]^2$ fallen mithin zusammen.

2) Setzt man

$$\begin{aligned}
(21) \quad & \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 = M, \\
& \frac{1}{r} \sum_{i=1}^r [z_i - z_{(r)}]^2 = M', \\
& \frac{r-1}{r^2} \sum_{i=1}^r \frac{1}{n_i(n_i-1)} \sum_{j=1}^{n_i} [x'_{i,j} - z_i]^2 = M'',
\end{aligned}$$

so hat man:

$$\begin{aligned}
E M' &= \frac{r-1}{r^2} \sum_{i=1}^r \frac{1}{n_i} \mu_2^{(i)} + \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2, \\
E M'' &= \frac{r-1}{r^2} \sum_{i=1}^r \frac{1}{n_i} \mu_2^{(i)}, \\
M &= E[M' - M''].
\end{aligned}$$

Man findet ferner:

$$\begin{aligned}
(21') \quad E[M' - E M']^2 &= \\
&= \frac{(r-1)^2}{r^4} \sum_{i=1}^r \frac{1}{n_i^2} [\mu_4^{(i)} - 3(\mu_2^{(i)})^2] + \frac{2(r-1)}{r^3} \sum_{i=1}^r \frac{1}{n_i^2} (\mu_2^{(i)})^2 - \\
&- \frac{2}{r^3} \sum_{i=1}^r \left[\frac{1}{n_i} \mu_2^{(i)} - \frac{1}{r} \sum_{j=1}^r \frac{1}{n_j} \mu_2^{(j)} \right]^2 + \frac{4}{r^2} \sum_{i=1}^r \frac{1}{n_i} \mu_2^{(i)} [m_1^{(i)} - m_{[1,r]}]^2 + \\
&+ \frac{4(r-1)}{r^3} \left\{ \sum_{i=1}^r \frac{1}{n_i^2} \mu_2^{(i)} m_1^{(i)} - m_{[1,r]} \sum_{i=1}^r \frac{1}{n_i^2} \mu_2^{(i)} \right\}, \\
E[M'' - E M'']^2 &= \frac{(r-1)^2}{r^4} \sum_{i=1}^r \frac{1}{n_i^2} [\mu_4^{(i)} - 3(\mu_2^{(i)})^2] + \\
&+ \frac{2(r-1)^2}{r^4} \sum_{i=1}^r \frac{1}{n_i^2 (n_i-1)} [\mu_2^{(i)}]^2,
\end{aligned}$$

$$E[M' - EM'] \cdot [M'' - EM''] = \frac{(r-1)^2}{r^4} \sum_{i=1}^r \frac{1}{n_i^2} [\mu_2^{(i)} - 3(\mu_2^{(i)})^2] + \\ + \frac{2(r-1)}{r^3} \left\{ \sum_{i=1}^r \frac{1}{n_i^2} \mu_2^{(i)} m_1^{(i)} - m_{[1,r]} \sum_{i=1}^r \frac{1}{n_i^2} \mu_2^{(i)} \right\}.$$

und hieraus:

$$(22) \quad E\{[M' - M''] - M\}^2 = \frac{2(r-1)}{r^4} \sum_{i=1}^r \frac{1}{n_i^2} [\mu_2^{(i)}]^2 + \\ + \frac{2(r-1)^2}{r^4} \sum_{i=1}^r \frac{1}{n_i(n_i-1)} [\mu_2^{(i)}]^2 - \frac{2}{r^3} \sum_{i=1}^r \left[\frac{1}{n_i} \mu_2^{(i)} - \frac{1}{r} \sum_{j=1}^r \frac{1}{n_j} \mu_2^{(j)} \right]^2 + \\ + \frac{4}{r^2} \sum_{i=1}^r \frac{1}{n_i} \mu_2^{(i)} [m_1^{(i)} - m_{[1,r]}]^2.$$

III.

1) Im Falle, wenn die Grössen $m_1^{(i)}$ untereinander gleich sind, findet man aus (21) und (22)

$$M = \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 = 0, \\ E[M' - M''] = 0, \\ (23) \quad E[M' - M'']^2 = \frac{2(r-1)}{r^4} \sum_{i=1}^r \frac{1}{n_i^2} [\mu_2^{(i)}]^2 + \\ + \frac{2(r-1)^2}{r^4} \sum_{i=1}^r \frac{1}{n_i(n_i-1)} [\mu_2^{(i)}]^2 - \frac{2}{r^3} \sum_{i=1}^r \left[\frac{1}{n_i} \mu_2^{(i)} - \frac{1}{r} \sum_{j=1}^r \frac{1}{n_j} \mu_2^{(j)} \right]^2.$$

Falls ausserdem die Grössen n_i unter einander gleich sind, hat man:

$$E[M' - M''] = \\ = \frac{1}{r} E \left\{ \sum_{i=1}^r [z_i - x_{(nr)}]^2 - \frac{r-1}{r} \frac{1}{n(n-1)} \sum_{i=1}^r \sum_{j=1}^n [x'_{i,j} - z_i]^2 \right\} = 0,$$

$$(24) \quad E[M' - M'']^2 = \frac{2(r-1)(nr-1)}{r^3 n^2 (n-1)} \frac{1}{r} \sum_{i=1}^r [\mu_2^{(i)}]^2 - \frac{2}{n^2 r^2} \frac{1}{r} \sum_{i=1}^r [\mu_2^{(i)} - \mu_{[2, r]}]^2.$$

Wenn ausser den Grössen $m_1^{(i)}$, auch die Grössen $m_2^{(i)}$ und folglich auch $\mu_2^{(i)}$ unter einander gleich sind, erhält man:

$$E[M' - M''] = 0,$$

$$(25) \quad E[M' - M'']^2 = \mu_2^2 \left\{ \frac{2(r-1)^2}{r^4} \sum_{i=1}^r \frac{1}{n_i(n_i-1)} - \frac{2}{r^4} \sum_{i=1}^r \frac{1}{n_i^2} + \frac{2}{r^4} \left[\sum_{i=1}^r \frac{1}{n_i} \right]^2 \right\}.$$

Falls ausserdem die Grössen n_i unter einander gleich sind, hat man:

$$(26) \quad E[M' - M'']^2 = \frac{2(r-1)(nr-1)}{r^3 n^2 (n-1)} \mu_2^2.$$

2) Der Fall, wo alle Variablen einem und demselben Verteilungsgesetze folgen, lässt sich auch in einer anderen Weise behandeln, die vor der obigen erhebliche Vorzüge hat. Man setze:

$$\frac{1}{r-1} \sum_{i=1}^r n_i [z_i - x_{(s)}]^2 = \mu'_2,$$

$$\frac{1}{s-1} \sum_{i=1}^s [x_i - x_{(s)}]^2 = \mu''_2.$$

Man hat:

$$E\mu'_2 = E\mu''_2 = \mu_2,$$

und folglich

$$E[\mu'_2 - \mu''_2] = 0.$$

Man findet ferner leicht

$$E\mu'_2 = \frac{r+1}{r-1}\mu_2^2 + \frac{1}{(r-1)^2} \left[\sum_{i=1}^r \frac{1}{n_i} - \frac{2r-1}{s} \right] \cdot [\mu_4 - 3\mu_2^2],$$

$$E\mu''_2 = \frac{s+1}{s-1}\mu_2^2 + \frac{1}{s} [\mu_4 - 3\mu_2^2]$$

und nach etwas beschwerlicheren Rechnungen

$$E\mu'_2\mu''_2 = \mu_2^2 + \frac{1}{s-1} [\mu_4 - \mu_2^2] - \frac{1}{s(s-1)} [\mu_4 - 3\mu_2^2],$$

woraus

$$(27) \quad E[\mu'_2 - \mu''_2]^2 = \frac{2}{r-1} \frac{s-r}{s-1} \mu_2^2 + \\ + \frac{1}{(r-1)^2} \left[\sum_{i=1}^r \frac{1}{n_i} - \frac{r^2}{s} \right] \cdot [\mu_4 - 3\mu_2^2]$$

folgt.

Sind die Grössen n_i untereinander gleich, so ist

$$\sum_{i=1}^r \frac{1}{n_i} - \frac{r^2}{s} = 0, \text{ und man erhält}$$

$$(28) \quad E[\mu'_2 - \mu''_2]^2 = \frac{2}{r-1} \frac{s-r}{s-1} \mu_2^2.$$

Wenn die Variable x dem GAUSS'schen Verteilungsgesetze folgt, ist $\mu_4 - 3\mu_2^2 = 0$, und man erhält gleichfalls

$$(29) \quad E[\mu'_2 - \mu''_2]^2 = \frac{2}{r-1} \frac{s-r}{s-1} \mu_2^2.$$

In diesem Falle ist also der mittlere Fehler von $\mu'_2 - \mu''_2$ von der Verteilung der Versuche auf die einzelnen Serien unabhängig.

Da $\sum_{i=1}^r \frac{1}{n_i} - \frac{r^2}{s} \geq 0$ ist, so wird das Vorzeichen des zweiten

Gliedes auf der rechten Seite von (27) durch das Vorzeichen der Differenz $\mu_4 - 3\mu_2^2$ bestimmt: ist $\mu_4 > 3\mu_2^2$, so wird

$E[\mu'_2 - \mu''_2]^2$ durch die ungleichmässige Verteilung der Versuche auf die einzelnen Serien vergrössert. Ist $\mu_4 < 3\mu_2^2$, so wird $E[\mu'_2 - \mu''_2]^2$ im Gegenteil hierdurch verringert.

3) In dem Falle, wenn alle Variablen einem und demselben Verteilungsgesetze folgen und die Zahlen n_i untereinander gleich sind, lässt sich ohne besondere Schwierigkeit die Gestalt des Verteilungsgesetzes von $\mu'_2 - \mu''_2$ für $n = \infty$ unter der Voraussetzung, dass $\left(\frac{n\mu_h}{[n\mu_2]^{h/2}}\right)_{n=\infty} = 0$ bei beliebig grossem h ist, feststellen.

Das Verteilungsgesetz der Grössen z_i nähert sich unter obiger Voraussetzung mit wachsendem n der GAUSS'schen Form an; man hat demnach

$$\begin{aligned} (E[\mu'_2]^k)_{n=\infty} &= \left(1 + \frac{2}{r-1}\right) \left(1 + \frac{4}{r-1}\right) \cdots \left(1 + \frac{2(k-1)}{r-1}\right) \mu_2^k, \\ (30) \quad (E[\mu'_2 - \mu_2]^k)_{n=\infty} &= \\ &= \mu_2^k \sum_{j=0}^k (-1)^j C_k^j \left(1 + \frac{2}{r-1}\right) \left(1 + \frac{4}{r-1}\right) \cdots \left(1 + \frac{2(k-j-1)}{r-1}\right). \end{aligned}$$

Anderseits hat man:

$$\begin{aligned} (E[\mu''_2]^k)_{n=\infty} &= \mu_2^k, \\ (E[\mu''_2 - \mu_2]^k)_{n=\infty} &= 0. \end{aligned}$$

In gleicher Weise, wie oben (zweite Abhandlung, S. 248) überzeugt man sich ferner leicht, dass bei beliebigen ganzzahligen positiven h und j

$$(E[\mu'_2 - \mu_2]^h \cdot [\mu''_2 - \mu_2]^j)_{n=\infty} = 0$$

und dass, folglich,

$$(31) \quad (E[\mu'_2 - \mu''_2]^k)_{n=\infty} = (E[\mu'_2 - \mu_2]^k)_{n=\infty}.$$

Das Verteilungsgesetz von $\mu'_2 - \mu''_2$ ist also bei $n = \infty$ dasselbe, wie das Verteilungsgesetz von $[\mu'_2 - \mu_2]$, d. i. das

PEARSON'sche asymmetrische Verteilungsgesetz, das dem Werte ∞ des PEARSON'schen Kriteriums entspricht¹ und dessen Gleichung auf die Form $y = kx^{\frac{r-3}{2}} e^{-\frac{r-1}{2\mu_2}x}$ gebracht werden kann, falls man den Nullpunkt des Coordinatensystems in den Punkt verlegt, welchem der Wert $y=0$ entspricht. Die Mode hat hierbei den Wert $\mu_2 \left[1 - \frac{2}{r-1}\right]$.

IV.

Die Bestimmung des Näherungswertes von $\frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2$

stellt auf die Tagesordnung die sich unmittelbar daran anschliessende Frage nach den Werten der höheren Momente der $m_1^{(i)}$ -Grössen. Ohne dieses komplizierte Problem in dessen ganzem Umfange hier aufzuwerfen, will ich nur in aller Kürze zeigen, wie sich in dem einfachsten Falle der gleichen Versuchszahlen der Wert von $\frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2$ auf Grundlage der empirischen Werte von $x'_{i,j}$ und z_i näherungsweise berechnen lässt.

Hält man an den Bezeichnungen des § I dieses Kapitels fest, so erhält man aus (2), indem man alle nr Versuche als ein Ganzes betrachtet:

$$\begin{aligned}
 (32) \quad E \frac{1}{nr} \sum_{i=1}^{nr} [x'_i - x_{(nr)}]^2 = & \\
 = \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2 + \frac{(nr-1)(nr-2)}{nr} \mu_{[3,r]} + & \\
 + \frac{3(nr-2)}{nr} \left[\frac{1}{r} \sum_{i=1}^r \mu_2^{(i)} m_1^{(i)} - \mu_{[2,r]} m_{[1,r]} \right]. &
 \end{aligned}$$

Wendet man anderseits die Formel (2) auf die z_i -Grössen an, wobei zu beachten ist, dass

¹ Cf. etwa W. PALIN ELDERTON, Frequency curves and correlation, p. 47, 68.

$$E[z_i - E z_i]^2 = \frac{1}{n} \mu_2, \quad E[z_i - E z_i]^3 = \frac{1}{n^2} \mu_3,$$

so findet man:

$$(33) \quad E \frac{1}{r} \sum_{i=1}^r [z_i - x_{(nr)}]^3 = \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^3 + \frac{(r-1)(r-2)}{n^2 r} \mu_{[3,r]} + \\ + \frac{3(r-2)}{nr} \left[\frac{1}{nr} \sum_{i=1}^{nr} \mu_2^{(i)} m_1^{(i)} - \mu_{[2,r]} m_{[1,r]} \right].$$

Aus

$$E \frac{1}{n} \sum_{j=1}^n [x'_{i,j} - z_i]^3 = \frac{(n-1)(n-2)}{n^2} \mu_3^{(i)}$$

findet man schliesslich:

$$(34) \quad E \frac{n}{r(n-1)(n-2)} \sum_{i=1}^r \sum_{j=1}^n [x'_{i,j} - z_i]^3 = \mu_{[3,r]}.$$

Man verfügt demnach über drei lineare Gleichungen, die nicht mehr, als drei, unbekannte Grössen enthalten. Aus (32) und (33) erhält man:

$$(35) \quad E \frac{1}{nr(nr-2)} \sum_{i=1}^{nr} [x'_i - x_{(nr)}]^3 - E \frac{1}{r(r-2)} \sum_{i=1}^r [z_i - x_{(nr)}]^3 = \\ = \frac{(n-1)(nr+r-1)}{n^2 r} \mu_{[3,r]} - \frac{r(n-1)}{(r-2)(nr-1)r} \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^3.$$

Setzt man in (35) den Wert von $\mu_{[3,r]}$ aus (34), so hat man schliesslich:

$$(36) \quad \frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^3 = \\ = E \left\{ \frac{(nr+r-1)(nr-1)(r-2)}{r^3 n(n-1)(n-2)} \sum_{i=1}^r \sum_{j=1}^n [x'_{i,j} - z_i]^3 + \right. \\ \left. + \frac{nr-2}{r^2(n-1)} \sum_{i=1}^r [z_i - x_{(nr)}]^3 - \frac{r-2}{r^2 n(n-1)} \sum_{i=1}^{nr} [x'_i - x_{(nr)}]^3 \right\}.$$

DRITTES KAPITEL.

I.

1) Man nehme nun in üblicher Weise (vgl. Erste Abhandlung, S. 210) an, dass die Variablen $x_1, x_2, \dots, x_r$ bloss Werte 1 und 0 annehmen können und bezeichne mit p_i die Wahrscheinlichkeit des Wertes 1 und mit $q_i (= 1 - p_i)$ die Wahrscheinlichkeit des Wertes 0 für die Variable x_i . Durch z_i wird dann die Häufigkeit des in bekannter Weise mit x_i zusammenhängenden Ereignisses A für die i -te Serie ausgedrückt und durch $x_{(s)}$ — die Häufigkeit von A für alle r Serien. Setzt man ferner

$$\frac{1}{r} \sum_{i=1}^r p_i = p_{(r)},$$

$$\frac{1}{s} \sum_{i=1}^s n_i p_i = p_{(s)},$$

so hat man

$$m_1^{(i)} = p_i, \quad m_{[1, r]} = p_{(r)}, \quad m_{[1, s]} = p_{(s)},$$

$$\mu_2^{(i)} = p_i q_i,$$

$$\sum_{i=1}^s [x'_i - x_{(s)}]^2 = \sum_{i=1}^s x_i^2 - s x_{(s)}^2 = \sum_{i=1}^s x_i - s x_{(s)}^2 = s x_{(s)} [1 - x_{(s)}]$$

und demnach

$$(37) \quad E \frac{1}{s} \sum_{i=1}^s [x'_i - x_{(s)}]^2 = E x_{(s)} [1 - x_{(s)}] =$$

$$= \frac{1}{s} \sum_{i=1}^s n_i [p_i - p_{(s)}]^2 + \frac{s-1}{s} \sum_{i=1}^s n_i p_i q_i,$$

$$(38) \quad E \frac{1}{r} \sum_{i=1}^r [z_i - z_{(r)}]^2 = \frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2 + \frac{r-1}{r^2} \sum_{i=1}^r \frac{1}{n_i} p_i q_i,^1$$

¹ Bleibt n_i für alle r Serien gleich, so erhält man aus der Formel (38) die von Prof. v. BORTKIEWICZ in der Abhandlung »Homogenität und

$$(39) \quad E \frac{1}{s} \sum_{i=1}^r n_i [z_i - x_{(s)}]^2 = \frac{1}{s} \sum_{i=1}^r n_i [p_i - p_{(s)}]^2 + \\ + \frac{1}{s} \sum_{i=1}^r p_i q_i - \frac{1}{s^2} \sum_{i=1}^r n_i p_i q_i,$$

$$(40) \quad E \frac{1}{r} \sum_{i=1}^r [z_i - x_{(s)}]^2 = \frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2 + \frac{1}{r} \sum_{i=1}^r \frac{1}{n_i} p_i q_i - \\ - \frac{2}{sr} \sum_{i=1}^r p_i q_i + \frac{1}{s^2} \sum_{i=1}^r n_i p_i q_i + 2p_{(s)}[p_{(s)} - p_{(r)}],$$

$$(41) \quad E \frac{1}{s} \sum_{i=1}^r n_i [z_i - z_{(r)}]^2 = \frac{1}{s} \sum_{i=1}^r n_i^2 [p_i - p_{(r)}]^2 + \\ + \frac{r-2}{sr} \sum_{i=1}^r p_i q_i + \frac{1}{r^2} \sum_{i=1}^r \frac{1}{n_i} p_i q_i.$$

Da anderseits

$$E z_i [1 - z_i] = p_i - p_i^2 - \frac{1}{n_i} p_i q_i = \frac{n_i - 1}{n_i} p_i q_i,$$

so ist

$$(42) \quad E \sum_{i=1}^r \frac{1}{n_i - 1} z_i [1 - z_i] = \sum_{i=1}^r \frac{1}{n_i} p_i q_i, \\ E \sum_{i=1}^r \frac{n_i^2}{n_i - 1} z_i [1 - z_i] = \sum_{i=1}^r n_i p_i q_i$$

und, folglich,

Stabilität in der Statistik» (Skandinavisk Aktuarietidskrift, 1918) auf S. 11 gegebene Formel (36). In gleicher Weise erhält man aus den Formeln (21') bzw. (10'), als einen Spezialfall, die von BORTKIEWICZ aufgestellte Formel (60) (ibid., S. 20), — oder vielmehr die genaue Formel, zu der BORTKIEWICZ gekommen wäre, falls er in den Formeln (54), (55) u. a. die $\frac{1}{s^2}$, bzw. $\frac{1}{s^3}$ enthaltenden Glieder nicht vernachlässigt hätte.

$$(43) \quad \frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2 = \\ = E \left\{ \frac{1}{r} \sum_{i=1}^r [z_i - z_{(r)}]^2 - \frac{r-1}{r^2} \sum_{i=1}^r \frac{1}{n_i-1} z_i [1-z_i] \right\},$$

$$(44) \quad \frac{1}{s} \sum_{i=1}^r n_i [p_i - p_{(s)}]^2 = \\ = E \left\{ x_{(s)} [1-x_{(s)}] - \frac{s-1}{s^2} \sum_{i=1}^r \frac{n_i^2}{n_i-1} z_i [1-z_i] \right\}.$$

2) Setzt man, wie oben (S. 13)

$$\begin{aligned} \frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2 &= M, \\ \frac{1}{r} \sum_{i=1}^r [z_i - z_{(r)}]^2 &= M', \\ \frac{r-1}{r^2} \sum_{i=1}^r \frac{1}{n_i-1} z_i [1-z_i] &= M'', \end{aligned}$$

so hat man

$$M = E[M' - M'']$$

und, laut (22),

$$(45) \quad E \{ [M' - M''] - M \}^2 = \frac{2(r-1)}{r^4} \sum_{i=1}^r \frac{1}{n_i^2} p_i^2 q_i^2 + \\ + \frac{2(r-1)^2}{r^4} \sum_{i=1}^r \frac{1}{n_i(n_i-1)} p_i^2 q_i^2 - \\ - \frac{2}{r^3} \sum_{i=1}^r \left[\frac{1}{n_i} p_i q_i - \frac{1}{r} \sum_{j=1}^r \frac{1}{n_j} p_j q_j \right]^2 + \frac{4}{r^2} \sum_{i=1}^r \frac{1}{n_i} p_i q_i [p_i - p_{(r)}]^2.$$

Wenn alle n_i unter einander gleich sind, erhält man:

$$\begin{aligned} \frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2 &= M = E \left\{ \frac{1}{r} \sum_{i=1}^r [z_i - z_{(r)}]^2 - \right. \\ &\quad \left. - \frac{r-1}{r^2(n-1)} \sum_{i=1}^r z_i [1 - z_i] \right\}, \\ (46) \end{aligned}$$

$$\begin{aligned} E \{ [M' - M''] - M \}^2 &= \frac{2(r-1)}{r^2 n} \left[\frac{1}{n} + \frac{r-1}{n-1} \right] \sum_{i=1}^r p_i^2 q_i^2 - \\ &\quad - \frac{2}{n^2 r^3} \sum_{i=1}^r \left[p_i q_i - \frac{1}{r} \sum_{j=1}^r p_j q_j \right]^2 + \frac{4}{n r^2} \sum_{i=1}^r p_i q_i [p_i - p_{(r)}]^2. \end{aligned}$$

¹ M' in den obigen Formeln entspricht der von Prof. v. BORTKIEWICZ in »Homogenität und Stabilität in der Statistik« (Skandin. Aktuarietidskrift, 1918) mit σ^2 bezeichneten Grösse; M'' entspricht dem BORTKIEWICZ'schen u^2 , M — dem BORTKIEWICZ'schen w^2 . Die Differenz $M' - M''$ ist somit dem BORTKIEWICZ'schen v^2 gleich, und $E \{ [M' - M''] - M \}^2 = \mathbb{M}^2(v^2)$. Was den Unterschied zwischen der obigen Formel (46) und der BORTKIEWICZ'schen Formel (103) (Homog. u. Stab., S. 27) anbelangt, so geht er, einerseits, darauf zurück, dass BORTKIEWICZ auf eine genaue Berechnung der in Betracht kommenden mathematischen Erwartungen verzichtet (cf. oben Anmerkung zu S. 20), und, andererseits, darauf, dass der in die Definition von v^2 (ibid., S. 26) eingehende empirische Wert u^2 bei der Berechnung von $\mathbb{M}^2(v^2)$ durch eine apriorische Grösse ersetzt wird. Berücksichtigt man die Identitäten

$$\frac{1}{r} \sum_{i=1}^r p_i^2 q_i^2 = p_{(r)}^2 q_{(r)}^2 + [1 - 6 p_{(r)} q_{(r)}] \varepsilon_2 + 2 [p_{(r)} - q_{(r)}] \varepsilon_3 + \varepsilon_4,$$

$$\frac{1}{r} \sum_{i=1}^r p_i q_i = p_{(r)} q_{(r)} - \varepsilon_2,$$

$$\frac{1}{r} \sum_{i=1}^r p_i q_i [p_i - p_{(r)}]^2 = p_{(r)} q_{(r)} \varepsilon_2 - [p_{(r)} - q_{(r)}] \varepsilon_3 - \varepsilon_4,$$

wobei $\varepsilon_k = \frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^k$ ist (cf. BORTKIEWICZ, Homog. und Stab., S. 21;

vgl. CHARLIER, Theorems of POISSON and LEXIS, p. 14, Ark. f. mat., astr. och fysik, 1911), so erhält man aus Formel (46) nach Weglassen der ε_2^2 , ε_3 und ε_4 enthaltenden Glieder mit der von BORTKIEWICZ angestrebten Annäherung

Wenn, anderseits, alle p_i untereinander gleich sind, erhält man:

$$M = 0,$$

$$(47) \quad E[M' - M'']^2 = p^2 q^2 \left\{ \frac{2(r-1)}{r^4} \sum_{i=1}^r \frac{1}{n_i^2} + \frac{2(r-1)^2}{r^4} \sum_{i=1}^r \frac{1}{n_i(n_i-1)} - \frac{2}{r^3} \sum_{i=1}^r \left[\frac{1}{n_i} - \frac{1}{r} \sum_{j=1}^r \frac{1}{n_j} \right]^2 \right\}.$$

Wenn, schliesslich, sowohl alle p_i , wie auch alle n_i gleich sind, erhält man:

$$M = 0,$$

$$(48) \quad E[M' - M'']^2 = \frac{2(r-1)(nr-1)}{r^3 n^2 (n-1)} p^2 q^2.$$

II.

Wenn alle p_i untereinander gleich sind, kann man auch von den Formeln des § III, 2) des zweiten Kapitels ausgehen. Man erhält dann:

$$(49) \quad u'_2 = \frac{1}{r-1} \sum_{i=1}^r n_i [z_i - x_{(s)}]^2,$$

$$u''_2 = \frac{s}{s-1} x_{(s)} [1 - x_{(s)}],$$

$$\begin{aligned} \mathfrak{M}^2(v^2) &= \frac{2}{r^2 n^2} \left\{ \frac{(r-1)(nr-1)}{r(n-1)} p^2 q^2 + 2 \frac{n}{n-1} (nr-r+5) p q \varepsilon_2 + \frac{n}{n-1} (r-2) \varepsilon_2 \right\} = \\ &= \frac{2}{r^2 n^2} \{ (r-1) p^2 q^2 + 2(nr+5) p q \varepsilon_2 + (r-2) \varepsilon_2 \} \text{ anstatt} \end{aligned}$$

$\mathfrak{M}^2(v^2) = \frac{2}{r^2 n^2} \{ (r-1) p^2 q^2 + 2nr p q \varepsilon_2 \}$ laut Formel (103): die Näherungsformel (103) darf also insoweit, als correct, gelten, inwiefern ε_2 derselben Grössenordnung, wie $\frac{1}{n}$, ist.

$$E[\mu'_2 - \mu''_2] = 0,$$

$$(50) \quad E[\mu'_2 - \mu''_2]^2 = \frac{2}{r-1} \frac{s-r}{s-1} p^2 q^2 + \\ + \frac{1}{(r-1)^2} \left[\sum_{i=1}^r \frac{1}{n_i} - \frac{r^2}{s} \right] pq[1-6pq].$$

Bei $p = q = \frac{1}{2}$ hat man demnach:

$$E[\mu'_2 - \mu''_2]^2 = \frac{1}{8(r-1)} \left\{ \frac{s-r}{s-1} - \frac{1}{r-1} \left[\sum_{i=1}^r \frac{1}{n_i} - \frac{r^2}{s} \right] \right\};$$

der mittlere Fehler von $\mu'_2 - \mu''_2$ ist also bei $p = \frac{1}{2}$ um so geringer, je ungleichmässiger sich die Versuche auf die einzelnen Serien verteilen.

Die ungleichmässige Verteilung der Versuche wirkt ermässigend auf $E[\mu'_2 - \mu''_2]^2$, solange $1 - 6pq < 0$ oder $\left(p - \frac{1}{2}\right)^2 < \frac{1}{12}$ bleibt. Bei $pq = \frac{1}{6}$ wird $E[\mu'_2 - \mu''_2]^2$ durch die Verteilung der Versuche auf die einzelnen Serien nicht beeinflusst. Bei $pq < \frac{1}{6}$ stellt sich $E[\mu'_2 - \mu''_2]^2$ um so grösser, je ungleichmässiger die Verteilung ist.

Sind die Grössen n_i untereinander gleich, so ergibt sich aus (49) und (50):

$$(51) \quad \mu'_2 = \frac{n}{r-1} \sum_{i=1}^r [z_i - x_{(nr)}]^2,$$

$$\mu''_2 = \frac{nr}{nr-1} x_{(nr)} [1 - x_{(nr)}],$$

$$(52) \quad E[\mu'_2 - \mu''_2]^2 = \frac{2}{r-1} \frac{r(n-1)}{nr-1} p^2 q^2,$$

und aus (31), bei nicht allzu kleinen p und q ,

$$(53) \quad (E[\mu'_2 - \mu''_2]^k)_{n=\infty} = \\ = p^k q^k \sum_{j=0}^k (-1)^j C_k^j \left(1 + \frac{2}{r-1}\right) \left(1 + \frac{4}{r-1}\right) \cdots \left(1 + \frac{2(k-j-1)}{r-1}\right).$$

Das Verteilungsgesetz von $\mu'_2 - \mu''_2$ bei $n = \infty$ und $\left(\frac{1}{pqn}\right)_{n=\infty} = 0$ hat also die Form (cf. oben S. 18):

$$y = k x^{\frac{r-3}{2}} e^{-\frac{r-1}{2pq} x}.$$

III.

Aus rechnerischen Gründen ist es oft zweckmässig, die Zahlen der Wiederholungen des Ereignisses A und nicht dessen Häufigkeiten den Berechnungen zu Grunde zu legen. Bezeichnet man mit w_i die Zahl der Wiederholungen von A bei den n_i Versuchen der i -ten Serie und mit w_s die Zahl der Wiederholungen von A im Laufe aller s Versuche, so hat man:

$$w_i = n_i z_i,$$

$$w_s = \sum_{i=1}^r w_i = s \mathcal{X}(s).$$

Sind alle n_i gleich, so erhält man aus (43) und (44):

$$(54) \quad \frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2 = \\ = E \left\{ \frac{1}{n^2 r} \sum_{i=1}^r \left[w_i - \frac{1}{r} w_{nr} \right]^2 - \frac{r-1}{n^2 r^2 (n-1)} \sum_{i=1}^r w_i [n - w_i] \right\} = \\ = E \left\{ \frac{1}{n^2 r^2} w_{nr} [nr - w_{nr}] - \frac{nr-1}{n^2 r^2 (n-1)} \sum_{i=1}^r w_i [n - w_i] \right\}.$$

Anderseits erhält man, wenn alle p_i gleich sind, aus (44)

$$(55) \quad E \left\{ \frac{1}{s^2} \left\{ w_s[s - w_s] - (s-1) \sum_{i=1}^r \frac{1}{n_i-1} w_i[n_i - w_i] \right\} \right\} = 0$$

und aus (49)

$$(56) \quad E \left\{ \frac{1}{r-1} \sum_{i=1}^r \frac{1}{n_i} \left[w_i - \frac{n_i}{s} w_s \right]^2 - \frac{1}{s(s-1)} w_s[s - w_s] \right\} = 0.$$

Sind sowohl alle n_i , wie alle p_i gleich, so hat man einerseits:

$$(57) \quad \begin{aligned} E \left\{ \frac{1}{r} \sum_{i=1}^r \left[w_i - \frac{1}{r} w_{nr} \right]^2 - \frac{r-1}{r} \frac{1}{n-1} \sum_{i=1}^r w_i[n - w_i] \right\} &= 0, \\ E \left\{ \frac{1}{r} \sum_{i=1}^r \left[w_i - \frac{1}{r} w_{nr} \right]^2 - \frac{r-1}{r} \frac{1}{n-1} \sum_{i=1}^r w_i[n - w_i] \right\}^2 &= \\ &= \frac{2(r-1)(nr-1)}{r^3(n-1)} (np)^2 q^2 \end{aligned}$$

und, andererseits,

$$(58) \quad \begin{aligned} E \left\{ \frac{1}{r-1} \sum_{i=1}^r \left[w_i - \frac{1}{r} w_{nr} \right]^2 - \frac{1}{r(nr-1)} w_{nr}[nr - w_{nr}] \right\} &= 0, \\ E \left\{ \frac{1}{r-1} \sum_{i=1}^r \left[w_i - \frac{1}{r} w_{nr} \right]^2 - \frac{1}{r(nr-1)} w_{nr}[nr - w_{nr}] \right\}^2 &= \\ &= \frac{2}{r-1} \frac{r(n-1)}{nr-1} (np)^2 q^2. \end{aligned}$$

Von besonderem Interesse ist der Fall, wo p so klein ist, dass w_{nr} im Vergleich zu nr vernachlässigt werden darf. Setzt man $p = \frac{m}{n}$ und lässt dann n unbeschränkt wachsen, so erhält man aus (57) und aus (58) übereinstimmend:

$$(59) \quad E \left\{ \frac{1}{r} \sum_{i=1}^r \left[w_i - \frac{1}{r} w_{nr} \right]^2 - \frac{1}{r} w_{nr} \right\} = 0,$$

$$(60) \quad E \left\{ \frac{1}{r-1} \sum_{i=1}^r \left[w_i - \frac{1}{r} w_{nr} \right]^2 - \frac{1}{r} w_{nr} \right\}^2 = \frac{2}{r-1} m^2. {}^1$$

In diesem Falle — d. i. im Bereiche des sogenannten »Gesetzes der kleinen Zahlen« — erhält man mithin Beziehungen, welche die Zahl n überhaupt nicht enthalten. Alle Berechnungen lassen sich ausführen, ohne dass man die genaue Grösse von n kennt: um die Formeln (59) und (60) anwenden zu dürfen, braucht man nur, berechtigt zu sein, anzunehmen, dass alle n_i untereinander gleich sind und dass n sehr gross ist im Vergleich zu der grössten der Zahlen w_i .

IV.

Bei der praktischen Verwertung der oben abgeleiteten Formeln wird man in der Regel die in denselben vorkommenden apriorischen Grössen durch empirische Werte ersetzen müssen. Am nächstliegenden ist gewiss, einfach überall $p_i = z_i$ und $q_i = 1 - z_i$ zu setzen, wie man es eben meistens zu tun pflegt. In der Praxis wird man sich wohl in der Regel an dieses Verfahren halten. Theoretisch ist dasselbe jedoch nicht einwandfrei. Man hat ja:

$$E z_i [1 - z_i] = \frac{n_i - 1}{n_i} p_i q_i,$$

woraus

$$p_i q_i = E \frac{n_i}{n_i - 1} z_i [1 - z_i], \text{ — also } p_i q_i \neq E z_i [1 - z_i] \text{ — folgt;}$$

bei kleinen Werten von n_i kann der der Annahme $p_i q_i = E z_i [1 - z_i]$ anhaftende Fehler recht erheblich sein.

Aus

$$E z_i^2 [1 - z_i]^2 = \frac{(n_i - 1)^2}{n_i^3} p_i q_i + \frac{(n_i - 1)(n_i - 2)(n_i - 3)}{n_i^3} p_i^2 q_i^2$$

¹ Cf. BORTKIEWICZ, Ueber die Zeitfolge zufälliger Ereignisse, S. 62 (Bull. Inst. Intern. de Stat., t. XX, Livr. 2).

findet man in ähnlicher Weise, bei $z_i \neq \frac{1}{n_i}$,

$$p_i^2 q_i^2 = \frac{n_i^3}{(n_i-1)(n_i-2)(n_i-3)} E \left\{ z_i^2 [1-z_i]^2 - \frac{n_i-1}{n_i^2} z_i [1-z_i] \right\}.$$

Hätte man, von $p_i q_i = \frac{n_i}{n_i-1} z_i [1-z_i]$ ausgehend, $p_i^2 q_i^2$ gleich $\frac{n_i^2 z_i^2 [1-z_i]^2}{(n_i-1)^2}$ gesetzt, so hätte man den Wert von $p_i^2 q_i^2$ überschätzt. Bei $p=q=\frac{1}{2}$ würde der relative Fehler bei nicht zu kleinem n_i nicht allzu schwer ins Gewicht fallen, aber bei stark von $\frac{1}{2}$ abweichenden Werten von p und q würde er selbst bei sehr grossem n_i recht erheblich sein können.

Der in (45) vorkommende Ausdruck

$$\frac{1}{r} \sum_{i=1}^r \left[\frac{1}{n_i} p_i q_i - \frac{1}{r} \sum_{j=1}^r \frac{1}{n_j} p_j q_j \right]^2$$

lässt sich in der Weise empirisch abschätzen, dass man von

$$\begin{aligned} \sum_{i=1}^r \left[\frac{1}{n_i} p_i q_i - \frac{1}{r} \sum_{j=1}^r \frac{1}{n_j} p_j q_j \right]^2 &= \sum_{i=1}^r \frac{1}{n_i^2} p_i^2 q_i^2 - \frac{1}{r} \left[\sum_{i=1}^r \frac{1}{n_i} p_i q_i \right]^2 = \\ &= \frac{r-1}{r} \sum_{i=1}^r \frac{1}{n_i^2} p_i^2 q_i^2 - \frac{2}{r} \sum_{i=1}^{r-1} \sum_{j=i+1}^r \frac{1}{n_i} p_i q_i \frac{1}{n_j} p_j q_j \end{aligned}$$

ausgeht und für $\frac{1}{n_i} p_i q_i$ als Näherungswert $\frac{1}{n_i-1} z_i [1-z_i]$, sowie für $\frac{1}{n_i^2} p_i^2 q_i^2$ den Wert $\frac{n_i}{(n_i-1)(n_i-2)(n_i-3)} \left\{ z_i^2 [1-z_i]^2 - \frac{n_i-1}{n_i} z_i [1-z_i] \right\}$ einsetzt.

Da ferner

$$\begin{aligned}
\sum_{i=1}^r \frac{1}{n_i} p_i q_i [p_i - p_{(r)}]^2 &= \\
&= \frac{(r-1)^2}{r^2} \sum_{i=1}^r \frac{1}{n_i} p_i^3 q_i - \frac{2(r-1)}{r^2} \sum_{i=1}^r \sum_{j \neq i} \frac{1}{n_i} p_i^2 q_i p_j + \\
&+ \frac{1}{r^2} \sum_{i=1}^r \sum_{j \neq i} \frac{1}{n_i} p_i q_i p_j^2 + \frac{1}{r^2} \sum_{i=1}^r \sum_{j \neq i} \sum_{h \neq i \neq j} \frac{1}{n_i} p_i q_i p_j p_h
\end{aligned}$$

ist, so braucht man hier bloss $q_i = 1 - p_i$ zu setzen und dann p durch z , p^2 durch $z^2 - \frac{1}{n-1} z[1-z]$ u. s. w. zu ersetzen, um einen adäquaten empirischen Näherungswert auch für diesen Ausdruck eventuell zu erhalten.

V.

An die Bestimmung des Näherungswertes von $\frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2$

lässt sich, wie oben (§ IV des zweiten Kapitels) die Frage nach den Werten der höheren Momente der p_i -Grössen anschliessen. Wenn die Versuchszahlen n_i untereinander gleich sind, erhält man leicht aus (36):

$$\begin{aligned}
(61) \quad \frac{1}{r} \sum_{i=1}^r [m_i^{(2)} - m_{[1,r]}]^2 &= \\
&= E \left\{ \frac{(nr + r - 1)(nr - 1)(r - 2)}{r^3(n - 1)(n - 2)} \sum_{i=1}^r z_i [1 - z_i] [1 - 2z_i] + \right. \\
&+ \left. \frac{nr - 2}{r^2(n - 1)} \sum_{i=1}^r [z_i - x_{(nr)}]^3 - \frac{r - 2}{r(n - 1)} x_{(nr)} [1 - x_{(nr)}] [1 - 2x_{(nr)}] \right\},
\end{aligned}$$

denn

$$\sum_{j=1}^n [x'_{i,j} - z_i]^3 = n z_i [1 - 3z_i + 2z_i^2] = n z_i [1 - z_i] [1 - 2z_i]$$

und

$$\sum_{i=1}^{nr} [x'_i - x_{(nr)}]^3 = nr x_{(nr)} [1 - x_{(nr)}] [1 - 2x_{(nr)}].$$

VIERTES KAPITEL.

I.

Die Berechnung der oben mit M bezeichneten (cf. S. 8, 13, 22) LEXIS'schen »wesentlichen Schwankungskomponente« (vgl. Erste Abhandlung, S. 214) kann ein mannigfaltiges statistisches Interesse bieten.

1) Als Quadrat des mittleren Fehlers der den empirischen Zahlen zu Grunde liegenden apriorischen Grössen $m_1^{(i)}$ bezw. p_i , besitzt M einen hervorragenden unmittelbaren Erkenntniswert. Durch die Berechnung von M lässt sich ein Urteil darüber bilden, inwieweit die Grössen $m_1^{(i)}$ bezw. p_i von den Merkmalen beeinflusst werden, auf welchen die Serienbildung beruht. Man fasse etwa die so oft behandelte, aber doch unentschiedene Frage nach dem Einflusse der Alterskombination der Eltern auf das Geschlecht der Nachkommenschaft ins Auge. Wie eifrig hat man seit HOFACKER und SADLER nach dem gesetzmässigen Zusammenhange zwischen der Sexualproportion der Geborenen und dem Alter der Eltern geforscht und wie ergebnislos sind bis jetzt alle derartige Forschungen geblieben! Kaum wähte ein Forscher eine gewisse Gesetzmässigkeit erfasst zu haben, als ein Anderer kam und an neuem empirischen Material überzeugend nachwies, dass von der eben entdeckten Regelmässigkeit keine Spur zu merken sei. Und daraufhin wiederholte sich dieselbe Geschichte in ewigem Turnus von neuem.¹ Diese Misserfolge wurzelten nicht zuletzt darin, dass man über geeignete Mittel zur Veredelung des empirischen Materials nicht verfügte und das hinter dem Schleier der zufälligen Schwankungen schimmernde Bild festzuhalten sich beeilte,

¹ Cf. etwa die Literaturübersicht bei E. ROLCKE, Einfluss des Alters der Eltern auf das Geschlecht der Kinder, S. 49—52 (Allg. Stat. Archiv, Bd. IX, 1915).

ohne vorher festgestellt zu haben, dass die beobachteten Unterschiede in den statistischen Häufigkeiten über die Grenzen des rein Zufälligen hinausgehen. Die erste Frage hätte stets sein sollen, ob überhaupt merkbare Unterschiede in den Wahrscheinlichkeiten der Knabengeburt zwischen den einzelnen Elterngruppen vorhanden sind. Erst wenn man Grund hat, diese Frage bejahend zu beantworten, lässt sich mit Aussicht auf Erfolg die weitere Frage stellen, für welche Gruppen die betreffende Wahrscheinlichkeit höher und für welche sie niedriger zu veranschlagen ist. Wird das Problem in dieser Weise zergliedert, so lässt sich die Vorfrage an der Hand der oben abgeleiteten Formeln stets insoweit lösen, dass man feststellt, ob Differenzen zwischen den Wahrscheinlichkeiten von der Grösse existieren, dass sie den durch den mittleren Fehler von M bestimmten Spielraum übersteigen.

Dasselbe Verfahren kann ferner dazu benützt werden, Gruppierungen nach verschiedenen Merkmalen unter einander zu vergleichen, wobei es sich also entscheiden lässt, wo die den Forscher interessierenden Differenzen der Wahrscheinlichkeiten grösser sind. Hätte man z. B. auf diese Weise festgestellt, ob Unterschiede im Alter der Väter oder Unterschiede im Alter der Mütter oder Altersunterschiede zwischen den Eltern die Wahrscheinlichkeit der Knabengeburt stärker beeinflussen, so wäre unser Wissen von den rätselhaften Vorgängen bei der Geschlechtsbestimmung nicht unwesentlich gefördert. Sehr lehrreich könnte sich auch der

Vergleich der Grösse von $\frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2$ für die Lebend-

geborenen, einerseits, und für die Gesamtheit der Geborenen (i. e. der Lebend- und Totgeborenen), anderseits, gestalten. Die praktische Verwertbarkeit dieses Forschungsverfahrens in dessen einfacherer Form, welche von der Annahme gleicher Versuchszahlen für alle Serien ausgeht, ist relativ gering, da man nur in Ausnahmefällen ein dieser Annahme nicht zu krass widersprechendes empirisches Material zur Verfügung bekommt. Durch die oben abgeleiteten allgemeinen Formeln wird dieses methodische Hindernis beseitigt, und es steht der praktischen Anwendung dieses vielverheissenden Verfahrens

nichts im Wege ausser der zeitraubenden Kalkulationsarbeit, die freilich von Einzelforschern, welche über kein Bureau-personal verfügen, in dem erforderlichen Umfange kaum bewältigt werden kann.¹

2) Die Berechnung von M lässt sich unter Umständen an die Ausführung der Ausgleichung mit Vorteil anschliessen.

¹ In den Tabellen der Berliner Statistik werden z. B. für die ehelich Geborenen die Zahlen der Knaben- und Mädchengeburten seit 1878 jahraus jahrein für mehr als 1500 Elterngruppen ausgewiesen. Ein Material, fürwahr, das allein, wenn es zweckentsprechend bearbeitet wäre, die Frage nach dem Vorhandensein eines Zusammenhanges zwischen der Sexualproportion der Geborenen und dem Alter der Eltern entscheiden könnte, das aber so gut, wie brach liegt und ein weiteres halbes Jahrhundert brach liegen wird, wenn nicht inzwischen statistische Forschungsinstitute entstehen, welche mit technischen Mitteln ausreichend ausgerüstet sind, um vor grösseren Aufgaben nicht zurückzuschrecken. Die Forscher, welche sich für das Problem der geschlechtsbestimmenden Faktoren interessieren, möchte ich an dieser Stelle vor der Heranziehung der Pariser Statistik ausdrücklich warnen. Die in den Pariser Tabellen enthaltenen Zahlen scheinen nämlich zum Teil nicht auf der Beobachtung zu beruhen, sondern mehr oder weniger frei erdichtet zu sein. Die Knabenquote schwankt in einer Weise, die allem widerspricht, was wir sonst beobachten: bald sind die Schwankungen ungeheuer gross, bald sind sie im Gegenteil unglaublich gering. Im Jahre 1898 treffen wir z. B. Gruppen mit keiner einzigen Mädchengeburt und 11, 13, 17 (zweimal), 26, 30, ja 32 Knabengeburten, eine Gruppe mit 25 Mädchengeburten und keiner Knabengeburt und zahlreiche Gruppen, wo das Missverhältnis zwar nicht so grell in die Augen schlägt, aber doch kaum glaubwürdiger ist (z. B. 558 Knaben und 268 Mädchen, 60 Knaben und 155 Mädchen, 99 Knaben und 205 Mädchen u. s. w., — cf. *Annuaire Statistique de la Ville de Paris* für 1898). Wenn wir anderseits die 8 Jahrgänge 1890/97 ins Auge fassen, so finden wir in den betr. Tabellen 147 Gruppen mit je drei Geburten, aber darunter keine einzige, wo alle drei geborene Mädchen wären, und bloss 8 Gruppen mit je drei Knaben; unter den Gruppen mit je 4, je 5 Geburten sind die extremen Fälle gleichfalls so gut, wie nicht vertreten. Betrachtet man mithin die Tabellen in ihrer Gesamtheit etwas aufmerksamer, so bleibt kein Zweifel übrig, dass die Zahlen — wenigstens teilweise — mit der Wirklichkeit nichts zu tun haben, sondern auf willkürlichen und nirgends, so weit ich sehe, erwähnten Manipulationen beruhen. Diese »Interpolationen« sind vermutlich von den mit der Aufbereitung beauftragten unteren Beamten auf eigene Rechnung und ganz planlos vorgenommen worden, wobei es offenbar zwei »Schulen« gegeben hat: die eine neigte zur gleichmässigen Verteilung der Geburten auf Knaben und Mädchen, wegen die andere die Extremen bevorzugte. Es ist, natürlich, möglich, — ja wahrscheinlich, — dass nicht alle Zahlen solchen unkontrollierbaren Ursprungs sind und dass einzelne Jahrgänge von einer derartigen Miss-handlung überhaupt verschont blieben. Solange aber der Sachverhalt vom Pariser Statistischen Bureau nicht voll aufgeklärt wird, wozu eine erneute sorgfältigere Aufbereitung des Materials erforderlich wäre, müssen diese Tabellen für die Wissenschaft, als nicht existierend, gelten.

Etwas verdächtig sind übrigens auch die betr. Berliner Tabellen für 1884, 1886 und 1891, — namentlich in den jüngeren Altersklassen der Väter (cf. die oben zitierte Abhandlung von ROLKE, S. 59 und 68); eine Nachprüfung wäre auch hier erwünscht.

Hat man etwa eine unausgeglichene Sterbetafel in der Form von, in einjährigen Intervallen auf einander folgenden empirischen Werten der Sterbewahrscheinlichkeiten, — $p'_1, p'_2, \dots p'_\omega$, — so besteht die Aufgabe der Ausgleichung darin, die ursprüngliche unregelmässig schwankende Reihe der p' -Werte durch eine andere, »glattere«, Reihe zu ersetzen unter der Voraussetzung, dass die unregelmässigen Schwankungen der empirischen Häufigkeiten $p'_1, p'_2, \dots p'_\omega$ Zufälligkeiten zu verdanken sind, dass aber die Reihe der, ihnen zu Grunde liegenden mathematischen Wahrscheinlichkeiten einen sprungfreien Verlauf aufweist, der eben den Einfluss des Alters auf die Sterblichkeit zum Ausdruck bringt.

Die nähere Bestimmung der p -Werte geschieht bekanntlich entweder graphisch oder »mechanisch«, unter stillschweigender Annahme eines bestimmten gesetzmässigen Zusammenhanges zwischen den beiden Variablen, oder, schliesslich, unter direkter Anlehnung an einen, mehr oder weniger plausiblen, analytischen Ausdruck für diese Annahme. In allen drei Fällen verfährt man ziemlich willkürlich und hat zwischen der Scylla eines allzu weit gehenden Schematismus', der die eigenartige Wellenbewegung der empirischen Zahlen durch eine allzu einfache Kurve nivelliert, und der Charybdis einer allzu getreuen Wiedergabe der ursprünglichen Zahlen, welche rein zufällige Schwankungen fortbestehen lässt, zu manövrieren. Gegen die Gefahr einer zu weit gehenden Ausgleichung sucht man sich gelegentlich, — namentlich bei der graphischen Ausgleichung — dadurch, zu schützen, dass man auf Grundlage der mittleren Fehler der einzelnen p' -Werte beiderseits von der empirischen Kurve eine Zone absteckt, innerhalb welcher die ausgeglichene Kurve der p -Werte liegen soll. Die Festlegung der Grenzen für die zulässigen Abweichungen der einzelnen p -Werte von den entsprechenden p' -Werten lässt sich ferner dadurch ergänzen, dass man in ähnlicher Weise den Spielraum vorschreibt, innerhalb dessen gewisse Funktionen der Differenzen $p' - p$ (z. B. $\sum_{i=1}^r (p'_i - p_i)^2$) sich halten sollen.

Hierdurch werden wichtige objective Stützpunkte gewonnen, um sowohl der Überausgleichung, wie der nicht weit genug gehenden Ausgleichung einigermaßen vorzubeugen: bleibt

etwa $\sum_{i=1}^r (p'_i - p_i)^2$ hinter ihrem erwartungsmässigen Werte mehr zurück, als es dem mittleren Fehler entspricht, so ist dies ein Zeichen dafür, dass die Zufälligkeiten noch nicht ausgeglichen sind.¹

Die Berechnung von M sowohl für die Gesamtheit der p -Werte, wie namentlich für einzelne Abschnitte der p -Kurve kann weitere ähnliche Stützpunkte gewähren für das Bestreben, die Willkür bei der Ausführung der Ausglei- chung durch objectiv vorgezeichnete Daten nach Möglichkeit einzudämmen. Wenn z. B. ein bestimmter Abschnitt der p -Kurve wellenlos verläuft, wogegen die betreffenden p' -Werte grössere Schwankungen aufweisen, so lässt sich durch den Vergleich von M mit dem mittleren Fehler von M entscheiden, ob dies nicht darauf zurückgeht, dass man eine reelle Welle auf der Kurve der wahren p -Werte durch die Wahl eines ungeeigneten Ausgleichungsverfahrens, als »zufällig«, mit »ausgeglichen« hat. Merkt man z. B. an einer ausgeglichenen Sterblichkeitstafel die sonst oft deutlich zu Tage tretende Abnahme der männlichen Sterblichkeit in den Jahren nach dem Militärdienst nicht, so lässt sich durch die Berechnung von M für einen passenden Altersspielraum nachweisen, ob das eben nicht auf einem »zuviel an Ausgleichung« beruht. Es wäre freilich hierzu erforderlich, über eine Reihe so dicht aufeinander folgender p' -Werte zu verfügen, dass der mittlere Fehler von M für den gewählten Altersspielraum sich hinreichend klein im Vergleich zu den fraglichen Unterschieden in der Sterblichkeit zeige.

¹ Dieses Verfahren wird gelegentlich auch in der Form angewandt, dass man bei der Ausgleichung, als Bedingung, aufstellt, dass der Wert

von $\sum_{i=1}^r (p'_i - p_i)^2$ der mathematischen Erwartung von $\sum_{i=1}^r (p'_i - p_i)^2$, unter

der Voraussetzung der gegenseitigen Unabhängigkeit der Versuche, genau gleich sei. Hierdurch wird der an sich richtige Gedanke, dass die Summe der quadrierten Differenzen zwischen p'_i und p_i , wenn diese Differenzen rein zufällig sind, gewisse Grenzen nach oben, wie nach unten, nicht überschreiten kann, in einer Weise überspannt, die sich wahrscheinlichkeitsrechnerisch nicht rechtfertigen lässt.

II.

1) Die LEXIS'sche Dispersionstheorie gipfelt bekanntlich in der Unterscheidung der normalen, der übernormalen und der unternormalen Dispersion (bezw. Stabilität). Als normal stabil, werden Reihen von statistischen Zahlen bezeichnet, deren Schwankungen um ihren Mittelwert den aus dem stochastischen Schema einer unveränderlichen Wahrscheinlichkeit (bezw. eines unveränderlichen Verteilungsgesetzes) bei gegenseitiger Unabhängigkeit aller Versuche sich ergebenden Erwartungen entsprechen. Als übernormal gilt die Dispersion in dem Falle, wenn die Schwankungen grösser sind, als es dem Schema der normalen Dispersion entsprechen würde. Sind hingegen die Schwankungen geringer, so ist die Dispersion unternormal (bezw. die Stabilität übernormal).

Als Mittel zur Unterscheidung des Charakters der Dispersion gilt für die Vertreter der DORMOY-LEXIS'schen Richtung die Grösse des Divergenzcoeffizienten: die Dispersion gilt für normal, wenn $Q^2 = 1$ (oder vielmehr, wenn die Differenz $Q^2 - 1$ sich innerhalb des Spielraums der zufälligen Schwankungen hält); ist $Q^2 > 1$, so ist die Dispersion übernormal; ist $Q^2 < 1$, so ist die Dispersion unternormal.

Wir haben oben gesehen, dass, wenn die Voraussetzungen der normalen Dispersion erfüllt sind, EQ^2 tatsächlich gleich 1 ist (zweite Abhandlung, erstes Kapitel). Dies reicht jedoch nicht, um die Verwertung der Grösse von Q^2 , als eines Kriteriums, voll zu rechtfertigen: es wäre hierzu erforderlich, dass die Bedingung $EQ^2 = 1$ nicht nur eine notwendige, sondern auch eine genügende wäre, d. h. dass sie *nur* in dem Falle erfüllt sein könnte, wenn alle Voraussetzungen der normalen Dispersion zuträfen. Dass die Sache sich anders verhält, haben wir bereits gesehen (erste Abhandlung, S. 208): nicht nur der zufällige empirische Wert von Q^2 , sondern auch EQ^2 kann gleich 1 sein, obgleich die einzelnen Versuche nicht unabhängig von einander sind. Man betrachte etwa das folgende stochastische Schema.

Aus r Urnen, die schwarze und weisse Kugeln enthalten, werden s Kugeln gezogen, wobei auf die i -te Urne n_i

Ziehungen entfallen und $\sum_{i=1}^r n_i = s$ ist. Die gezogenen Kugeln werden in die Urne *nicht* zurückgelegt. Die Zahl der in der i -ten Urne vor dem Beginn der Ziehungen enthaltenen schwarzen Kugeln sei a_i , die der weissen — b_i . Man bezeichne mit z_i die Quote der schwarzen Kugeln unter den n_i aus der i -ten Urne gezogenen Kugeln, mit $x_{(s)}$ — die Gesamtquote der schwarzen Kugeln unter allen s gezogenen Kugeln und setze

$$\frac{a_i}{a_i + b_i} = p_i, \quad 1 - p_i = q_i.$$

Betrachtet man alle s Ziehungen, als ein Ganzes, und setzt in den Formeln des § I des ersten Kapitels der ersten Abhandlung

$$m_1^{(i)} = p_i, \quad m_2^{(i)} = p_i, \quad \mu_3^{(i)} = p_i q_i$$

$E x'_{h_1} x'_{h_2} - m_1^{(h_1)} m_1^{(h_2)} = 0$, wenn h_1 und h_2 Ziehungen aus verschiedenen Urnen bezeichnen,

$E x'_{h_1} x'_{h_2} - m_1^{(h_1)} m_1^{(h_2)} = -\frac{1}{a_i + b_i - 1} p_i q_i$, wenn sowohl h_1 , wie h_2 , Ziehungen aus der i -ten Urne bezeichnen (vgl. Formel (16) der ersten Abhandlung), so erhält man:

$$E x_{(s)} = \frac{1}{s} \sum_{i=1}^r n_i p_i = p_{(s)},$$

$$(62) \quad E x_{(s)}^2 = p_{(s)}^2 + \frac{1}{s^2} \sum_{i=1}^r n_i \frac{a_i + b_i - n_i}{a_i + b_i - 1} p_i q_i,$$

$$(63) \quad E \frac{s}{s-1} x_{(s)} [1 - x_{(s)}] = \frac{1}{s} \sum_{i=1}^r n_i p_i q_i + \\ + \frac{1}{s(s-1)} \sum_{i=1}^r \frac{n_i(n_i-1)}{a_i + b_i - 1} p_i q_i + \frac{1}{s-1} \sum_{i=1}^r n_i [p_i - p_{(s)}]^2.$$

Sind alle p_i untereinander gleich, so hat man:

$$(64) \quad E \frac{s}{s-1} x_{(s)} [1 - x_{(s)}] = p q \left\{ 1 + \frac{1}{s(s-1)} \sum_{i=1}^r \frac{n_i(n_i-1)}{a_i + b_i - 1} \right\}.$$

Setzt man anderseits in den Formeln des § I des ersten Kapitels der ersten Abhandlung

$$n_1^{(i)} = E z_i = p_i, \quad m_2^{(i)} = E z_i^2 = \frac{a_i(a_i-1)}{(a_i+b_i)(a_i+b_i-1)} + \\ + \frac{1}{n_i} \frac{a_i b_i}{(a_i+b_i)(a_i+b_i-1)}, \quad \mu_2^{(i)} = p_i q_i \frac{1}{n_i} \frac{a_i+b_i-n_i}{a_i+b_i-1},$$

so findet man:

$$(65) \quad E \frac{1}{r-1} \sum_{i=1}^r [z_i - z_{(r)}]^2 = \\ = \frac{1}{r-1} \sum_{i=1}^r [p_i - p_{(r)}]^2 + \frac{1}{r} \sum_{i=1}^r p_i q_i \frac{1}{n_i} \frac{a_i+b_i-n_i}{a_i+b_i-1}.$$

Wenn alle p_i unter einander gleich sind, folgt hieraus:

$$(66) \quad E \frac{1}{r-1} \sum_{i=1}^r [z_i - z_{(r)}]^2 = p q \left\{ \frac{1}{r} \sum_{i=1}^r \frac{1}{n_i} - \frac{1}{r} \sum_{i=1}^r \frac{n_i-1}{n_i} \frac{1}{a_i+b_i-1} \right\}.$$

In gleicher Weise erhält man:

$$(67) \quad E \frac{1}{r-1} \sum_{i=1}^r n_i [z_i - x_{(s)}]^2 = \frac{1}{r-1} \sum_{i=1}^r n_i E z_i^2 - \frac{s}{r-1} E x_{(s)}^2 = \\ = \frac{1}{r-1} \sum_{i=1}^r n_i [p_i - p_{(s)}]^2 + \frac{1}{r-1} \sum_{i=1}^r \frac{s-n_i}{s} \frac{a_i+b_i-n_i}{a_i+b_i-1} p_i q_i.$$

Sind alle p_i untereinander gleich, so hat man:

$$(68) \quad E \frac{1}{r-1} \sum_{i=1}^r n_i [z_i - x_{(s)}]^2 = p q \left\{ 1 - \frac{1}{r-1} \sum_{i=1}^r \frac{s-n_i}{s} \frac{n_i-1}{a_i+b_i-1} \right\}.$$

Wenn also die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen, vor dem Beginn des Experimentes für alle r Urnen die gleiche ist, so ist:

$$E \frac{s}{s-1} x_{(s)} [1 - x_{(s)}] > pq > E \frac{1}{r-1} \sum_{i=1}^r n_i [z_i - x_{(s)}]^2,$$

$$E \frac{s}{s-1} x_{(s)} [1 - x_{(s)}] > pq > E \frac{r}{(r-1) \sum_{i=1}^r \frac{1}{n_i}} \sum_{i=1}^r [z_i - z_{(r)}]^2.$$

Der Divergenzcoefficient würde demnach sich < 1 stellen und die Dispersion, als unternormal, kennzeichnen.

Wenn aber die Wahrscheinlichkeiten p_i unter einander nicht gleich sind, so ist wohl möglich, dass $E \frac{s}{s-1} x_{(s)} [1 - x_{(s)}]$ entweder $E \frac{1}{r-1} \sum_{i=1}^r n_i [z_i - x_{(s)}]^2$ oder $E \frac{r}{(r-1) \sum_{i=1}^r \frac{1}{n_i}} \sum_{i=1}^r [z_i - z_{(r)}]^2$

gleich ist. Im Falle, wenn alle n_i unter einander gleich sind, wäre z. B. hierzu bloss erforderlich, dass

$$\frac{1}{r} \sum_{i=1}^r \frac{1}{a_i + b_i - 1} p_i q_i = \frac{r}{r-1} \sum_{i=1}^r [p_i - p_{(r)}]^2$$

sei. Der Divergenzcoefficient würde dann gleich 1 sein, und die Dispersion wäre auf Grund dessen für normal zu erklären, obgleich keine der wesentlichen Voraussetzungen der normalen Dispersion erfüllt ist: die Wahrscheinlichkeiten p_i sind verschieden, und die einzelnen Versuche sind nicht unabhängig von einander. Der Umstand, dass $Q^2 = 1$ ist, bietet also an sich noch keine Bürgschaft dafür, dass die Voraussetzungen der normalen Dispersion tatsächlich erfüllt sind.

Abgesehen von diesem prinzipiellen Einwand, lässt sich gegen die Wahl der Divergenzcoefficienten, als Kriteriums bei der Unterscheidung des Charakters der Dispersion, die Unsicherheit der Bestimmung des mittleren Fehlers von Q^2 anführen, die jedoch in der Regel kaum allzu schwer ins Gewicht fallen kann, da es sich gezeigt hat, dass $\frac{2}{r-1}$ die Grenze darstellt, zu welcher $E[Q^2 - 1]^2$ selbst bei ungleichmässiger Verteilung der Versuche auf die Serien mit wach-

sender Zahl der Versuche strebt unter Voraussetzungen, deren Erfüllung unschwer zu kontrollieren ist (vgl. zweite Abhandlung, zweites Kapitel, § V).

2) An Stelle des Quotienten der Ausdrücke, welche im Zähler und im Nenner des Divergenzcoeffizienten stehen, hätte man ebenso gut deren Differenz, als Kriterium, verwenden können, und die Dispersion für normal gelten lassen, falls diese Differenz gleich 0 ist, — für übernormal, falls sie > 0 ist, — für unternormal, falls sie < 0 ist.

Versucht man die Gründe zu rekonstruieren, welche sowohl DORMOY, wie LEXIS bewogen haben mögen, das erste von den beiden gleich nahe liegenden Verfahren zu bevorzugen, so dürften sie wohl hauptsächlich darauf zurückgehen, dass die annähernde Gleichheit der beiden Ausdrücke in den Abweichungen des Quotienten von 1 drastischer zum Vorschein kommt, als in den Abweichungen der Differenz von 0, — ein Vorzug, der in den 70-er Jahren des XIX. Jahrhunderts von den Vertretern des wahrscheinlichkeitstheoretischen Standpunktes in der Statistik gewiss nicht gering zu veranschlagen war angesichts der Nachklänge des kaum abgeschlossenen erbitterten Kampfes gegen die Uebertreibungen des Queteletismus. Als weniger wichtiges Nebemotiv war vermutlich der Wunsch im Spiel, im Divergenzcoeffizienten nicht nur ein Kriterium, sondern auch ein rationelles Mass für die Abweichung der Dispersion von der »Norm« zu erhalten.

Keinem dieser Gründe kann gegenwärtig ein grösseres Gewicht beigelegt werden. Die stochastische Orientierung der statistischen Theorie bedarf keiner künstlichen Stütze mehr, und ausserdem besteht kein Unterschied mehr zwischen den beiden Verfahren in Bezug auf die Leichtfassbarkeit ihrer Ergebnisse, seit man die Notwendigkeit eingesehen hat, die Abweichungen in beiden Fällen durch die betreffenden mittleren Fehler zu dividieren. Was aber die Idee anbelangt, den Divergenzcoeffizienten als Mass für die Abweichung der Dispersion von der »Norm« zu verwerten, so hat LEXIS selbst überzeugend nachgewiesen, dass zu diesem Zwecke der Wert von Q^2 ganz ungeeignet erscheint und im Falle der übernormalen Dispersion durch die Berechnung

der »wesentlichen Schwankungscomponente« ersetzt werden soll.

Es erscheint demnach die Berechnung von Q^2 gegenwärtig, als ein veraltetes Verfahren, das nur aus Rücksicht auf eingebürgerte Gewohnheiten, sowie gelegentlich zu Popularisierungszwecken gebraucht zu werden verdient. Zu Forschungszwecken bedient man sich zweckmässiger eines anderen Verfahrens. Man soll zunächst auf Grund der Formeln des § III, 2) des zweiten Kapitels (bzw. des § II des dritten Kapitels) prüfen, inwieweit die Voraussetzungen der normalen Dispersion erfüllt zu sein scheinen. Zeigt sich hierbei, dass die Dispersion ausgesprochen übernormal ist, so soll man mit Hilfe der Formeln (21) und (22) (bzw. (45)) die Grösse M und ihren mittleren Fehler berechnen.¹ Ist die Dispersion hingegen ausgesprochen unternormal, so lässt man es sein, da für den Fall der unternormalen Dispersion kein rationelles Mass vorliegt, welches ein allgemeines Interesse beanspruchen dürfte. Sieht, schliesslich, die Dispersion normal aus, so lässt sich eventuell ein Versuch machen, aufzuklären, inwieweit von dem äusseren Bilde einer normalen Dispersion auf die Erfüllung aller Voraussetzungen der normalen Dispersion geschlossen werden darf (cf. unten § IV).

III.

1) Die lebhafte Polemik, welche durch die BORTKIEWICZ'sche Untersuchung über »das Gesetz der kleinen Zahlen« ausgelöst worden ist, verleiht ein spezifisches Interesse der Frage, inwieweit bzw. unter welchen Voraussetzungen Schlüsse in Bezug auf den Charakter der Dispersion gezogen werden können, wenn man bloss über die Zahlen der Wiederholungen eines Ereignisses für eine Reihe von Versuchsserien verfügt, ohne die Zahlen der Versuche für die einzelnen Serien zu kennen.

¹ Da die allgemeinen Formeln (21) und (22) den Fall der normalen Dispersion mitumfassen, so kann man natürlich auch in der Weise vorgehen, dass man ohne weiteres die Grösse M berechnet und auf normale Dispersion schliesst, falls ihr Wert hinreichend gering im Vergleich zu ihrem mittleren Fehler ist. Der Umweg über die Formeln des § III, 2) des zweiten und namentlich des § II des dritten Kapitels hat jedoch den Vorteil, dass die Rechnungen sich einfacher gestalten, und dass solche Gruppen hinzugezogen werden können, die aus je einem Versuche bestehen.

Ob die Annahmen einer normalen Dispersion und gleich grosser Versuchszahlen für alle Serien mit der Erfahrung übereinstimmen, lässt sich in folgender Weise prüfen. Man bezeichne, wie oben (S. 26), die Zahl der Wiederholungen des Ereignisses für die i -te Serie mit w_i , die unbekannte Wahrscheinlichkeit des Ereignisses mit p , die unbekannte, aber für alle Serien gleich grosse Zahl der Versuche mit n , und setze $\frac{1}{r} \sum_{i=1}^r w_i = w_{(r)}$. Unter der Voraussetzung normaler Dispersion erhält man aus

$$E w_i = n p, \quad E w_i^2 = n p q + n^2 p^2$$

genau

$$n p = E w_{(r)},$$

$$n p q = E \frac{1}{r} \sum_{i=1}^r [w_i - n p]^2 = E \frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2$$

und näherungsweise

$$n p = w_{(r)},$$

$$n p q = \frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2,$$

woraus:

$$(69) \quad \begin{aligned} q = 1 - p &= \frac{\frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2}{w_{(r)}}, \\ p &= \frac{w_{(r)} - \frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2}{w_{(r)}}, \\ n &= \frac{w_{(r)}^2}{w_{(r)} - \frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2} \end{aligned}$$

folgt.¹

¹ Da man bei der obigen Bestimmung der Werte von p , q und n von

Erscheinen nun die auf diese Weise gefundenen Werte von p , q oder n aus irgend welchen Gründen unmöglich, so ist man gezwungen, die Annahmen, von welchen man ausgegangen ist, fallen zu lassen. Dies ist z. B. der Fall, wenn einer der Werte negativ ausfällt und die Abweichung von 0 zu gross ist, um auf zufällige Schwankungen zurückgeführt zu werden.

Dieses, von K. PEARSON und dessen Schülerin L. WHITAKER¹ in Vorschlag gebrachte Verfahren lässt sich systematisch ausbauen. Unter den Voraussetzungen der normalen Dispersion und gleicher Versuchszahlen kann man nämlich bei der Bestimmung der unbekannten Werte von p und von n nicht nur von den erwartungsmässigen Werten der ersten und der zweiten Potenzen von w_i ausgehen, sondern auch analoge Beziehungen für höhere Potenzen zu Grunde legen, d. h.

$$E \frac{1}{r} \sum_{i=1}^r w_i^3 = n^3 p^3 + 3n^2 p^2 q^3 + n p q (q - p),$$

$$E \frac{1}{r} \sum_{i=1}^r w_i^4 = n^4 p^4 + 6n^3 p^3 q + n^2 p^3 q (7 - 11p) + n p q (1 - 6p q)$$

u. s. w.

Man erhält auf diese Weise beliebig viele Gleichungen mit nur zwei Unbekannten: p und n . Trifft man hierbei

der Annahme der normalen Dispersion ausgeht, so hat man identisch

$$Q^2 = \frac{\frac{1}{r-1} \sum_{i=1}^r [w_i - w(r)]^2}{n p q} = 1. \text{ Diese Identität bleibt auch in dem Falle}$$

bestehen, wenn man, unter der Voraussetzung, dass $(np)_{n=\infty}$ einen endlichen Wert m annimmt, n unbegrenzt wachsen lässt. Hierbei bleibt Q^2 fortwährend identisch gleich 1, wogegen q der Einheit, als einem Grenzwert zustrebt, so dass, bei $n = \infty$, $q = 1 = Q^2$ wird; solange aber dieser Grenzwert von q nicht erreicht wird, bleibt $q < 1$ und mithin $Q^2 > q$. Ein innerer Widerspruch der Art, wie ihn Prof. v. BORTKIEWICZ dem obigen Verfahren vorwirft (Realismus und Formalismus in der math. Statistik, S. 230), haftet somit gar nicht dem Wesen des Verfahrens an.

¹ Cf. L. WHITAKER, On the Poisson's law of small numbers (Biometrika, vol. X); K. PEARSON, Skew variations in homogenous material, p. 347 (Phil. Trans., A, vol. 186).

auf keine Discrepanzen, die über die Grenzen des Zufälligen hinausgehen, so dürfen die Annahmen, von welchen man ausgegangen ist, als durch die Erfahrung bestätigt gelten. Sonst muss man sie, als nicht stichhaltig, fallen lassen.

Hat man Grund anzunehmen, dass die Wahrscheinlichkeit p so gering ist, dass w_i im Vergleich zu n ohne grösseren Fehler vernachlässigt werden darf, so lässt sich auch in einer einfacheren Weise prüfen, ob die Voraussetzungen zutreffend sind. Man kann nämlich das übliche DORMOX-LEXIS'sche Kriterium anwenden, indem man von den Formeln (59) und (60) ausgeht bzw. den Wert des Divergenzcoefficienten berechnet, der sich in diesem Falle feststellen lässt, ohne dass man die Zahl n zu kennen braucht. Dies ist das von Prof. v. BORTKIEWICZ für den Geltungsbereich des »Gesetzes der kleinen Zahlen« bevorzugte Verfahren. Dieses Verfahren besitzt den unbestreitbaren Vorteil einer grösseren Einfachheit. Es darf aber nicht übersehen werden, dass dasselbe in der Praxis stets, als ein Näherungsverfahren, auftritt, dessen Berechtigung auf der an ∞ grenzen sollen den Grösse der unbekannten Versuchszahl n beruht. Das PEARSON-WHITAKER'sche Verfahren hat hingegen den Vorzug, an keine weiteren Annahmen in Bezug auf die Grösse der unbekannten p und n gebunden und insoweit allgemeiner und prinzipiell strenger zu sein. In den Fällen, wo man a priori nicht sicher sein kann, dass n sehr gross ist, erscheint eine Nachprüfung der auf Grund von (59) und (60) erzielten Ergebnisse an der Hand des PEARSON-WHITAKER'schen Verfahrens keineswegs überflüssig: überzeugt man sich hierbei, dass n nicht gross sein kann, so soll man auf die Anwendung der Formeln (59) und (60) lieber verzichten.

2) Zeigt nun das eine oder das andere der obigen Verfahren, dass die ihnen zu Grunde liegenden Annahmen nicht aufrecht erhalten werden können, so entsteht die Frage, welche unter den Annahmen der Wirklichkeit widersprechen, — namentlich, ob die Unstimmigkeiten nicht allein darauf zurückzuführen sind, dass die Versuchszahlen der einzelnen Serien verschieden sind.

Bezeichnet man mit n_i die Zahl der Versuche der i -ten Serie, so hat man, wenn die Voraussetzungen der normalen Dispersion erfüllt sind,

$$Ew_i = pn_i,$$

$$Ew_i^2 = p^2 n_i^2 + pq n_i.$$

Setzt man nun in der Formel (1) des ersten Kapitels

$$N = r, \quad \frac{1}{r} \sum_{i=1}^r w_i = w_{(r)}, \quad \frac{1}{r} \sum_{i=1}^r n_i = \frac{s}{r} = n_0,$$

$$m_1^{(i)} = pn_i, \quad m_2^{(i)} = p^2 n_i^2 + pq n_i, \quad \mu_2^{(i)} = pq n_i,$$

$$m_{[1,r]} = \frac{1}{r} \sum_{i=1}^r pn_i = p \frac{1}{r} \sum_{i=1}^r n_i = p \frac{s}{r} = pn_0,$$

$$\mu_{[2,r]} = \frac{1}{r} \sum_{i=1}^r pq n_i = pq \frac{s}{r} = pq n_0,$$

so erhält man genau

$$Ew_{(r)} = E \frac{1}{r} \sum_{i=1}^r w_i = pn_0 = \frac{1}{r} (ps),$$

$$(70) \quad E \frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2 = pq n_0 + \frac{r}{r-1} p^2 \frac{1}{r} \sum_{i=1}^r [n_i - n_0]^2 =$$

$$= \frac{1}{r} (ps)q + \frac{1}{r(r-1)} (ps)^2 \frac{1}{r} \sum_{i=1}^r \left[\frac{n_i}{n_0} - 1 \right]^2$$

und näherungsweise

$$pn_0 = w_{(r)},$$

$$pq n_0 + \frac{r}{r-1} p^2 \frac{1}{r} \sum_{i=1}^r [n_i - n_0]^2 = \frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2.$$

Man verfügt also über zwei Gleichungen mit drei Unbekannten — p , n_0 und $\frac{1}{r} \sum_{i=1}^r [n_i - n_0]^2$; durch Heranziehen der

höheren Momente lässt sich die Lage gleichfalls nicht verbessern, da die Zahl der neu hinzukommenden Gleichungen nicht schneller wächst, als die Zahl der Unbekannten.

Ist jedoch p so klein, dass man, für $s = \infty$, $p_r^s = m$ und $q = 1$ setzen darf, so hat man näherungsweise:

$$(71) \quad \frac{1}{r} \sum_{i=1}^r w_i = w_{(r)} = m,$$

$$(72) \quad \frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2 = m + m^2 \frac{1}{r-1} \sum_{i=1}^r \left[\frac{n_i}{n_0} - 1 \right]^2,$$

woraus

$$\frac{1}{r} \sum_{i=1}^r \left[\frac{n_i}{n_0} - 1 \right]^2 = \frac{\frac{1}{r} \sum_{i=1}^r [w_i - w_{(r)}]^2 - \frac{(r-1)}{r} w_{(r)}^2}{w_{(r)}^2}$$

folgt. Im Geltungsbereiche des »Gesetzes der kleinen Zahlen« kann man somit auf diese Weise einen ungefähren Begriff davon gewinnen, wie gross die Unterschiede zwischen den Versuchszahlen sein müssen, damit die Dispersion für normal gehalten werden könne.¹

¹ Aus (71) und (72) ergibt sich:

$$(A) \quad \frac{\frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2 - w_{(r)}^2}{\sqrt{\frac{2}{r-1} \cdot m}} = m \sqrt{\frac{1}{2(r-1)} \sum_{i=1}^r \left[\frac{n_i}{n_0} - 1 \right]^2}$$

In ähnlicher Weise findet man aus (54), wenn n sehr gross ist und w_i im Vergleich zu n vernachlässigt werden darf:

$$\frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2 = \frac{1}{n^2 r} \sum_{i=1}^r [w_i - w_{(r)}]^2 - \frac{r-1}{n^2 r} \frac{n}{n-1} w_{(r)}^2$$

und

$$(B) \quad \frac{\frac{1}{r-1} \sum_{i=1}^r [w_i - w_{(r)}]^2 - w_{(r)}^2}{\sqrt{\frac{2}{r-1} \cdot w_{(r)}}} = w_{(r)} \sqrt{\frac{1}{2(r-1)} \sum_{i=1}^r \left[\frac{p_i}{p_{(r)}} - 1 \right]^2}$$

3) Ist die Zahl der Serien — r — hinreichend gross, so lässt sich eventuell ein tieferer Einblick in die Dispersionsverhältnisse dadurch gewinnen, dass man die Serien zu je zwei (bezw. je drei u. s. w.) in zufälliger Weise zusammenschlägt. Bezeichnet man mit w'_j die Zahl der Wiederholungen des Ereignisses für die j -te der neuen $\frac{r}{2}$ Serien und mit n'_j die Zahl der betr. Versuche, so hat man

$$E \sum_{j=1}^{r/2} [n'_j - n'_0]^2 = E \sum_{i=1}^r [n_i - n_0]^2$$

und erhält eine dritte approximative Gleichung, welche keine neuen Unbekannten enthält, nämlich:

$$\frac{1}{r-2} \sum_{j=1}^{r/2} \left[w'_j - \frac{2}{r} \sum_{j=1}^{r/2} w'_j \right]^2 = p q n_0 + \frac{r}{r-2} p^2 \frac{1}{r} \sum_{i=1}^r [n_i - n_0]^2.$$

Wenn die Dispersion normal ist, d. i. alle p_i unter einander gleich und die Einzelversuche von einander unabhängig sind, und wenn ausserdem die n_i Werte gleich sind, ist die mathematische Erwartung der Differenz

$\frac{1}{r-1} \sum_{i=1}^r [w_i - w(r)]^2 - w(r)$ gleich 0; durch den Quotienten derselben in

ihren mittleren Fehler, der gleich $m \sqrt{\frac{2}{r-1}}$, bzw. $w(r) \sqrt{\frac{2}{r-1}}$, gesetzt

wird, wird die Wahrscheinlichkeit gemessen, dass die Abweichung der obigen Differenz von 0 den Spielraum der zufälligen Schwankungen übersteigt. Die Formeln (A) und (B) zeigen, dass in dem Falle, wenn die p_i -Werte oder die n_i -Werte untereinander verschieden sind, dieser Quotient um so geringer ausfällt, je geringer m , bzw. $w(r)$, ist. Durch das übliche Forschungsverfahren werden somit Abweichungen im Verhalten der empirischen Zahlenreihe von den Erwartungen, welche sich aus den zu kontrollierenden Annahmen ergeben, um so mehr verdeckt, je geringer die durchschnittliche Wiederholungszahl des Ereignisses ist. Gleich grosse Un-

gleichheiten zwischen den $\frac{p_i}{p(r)}$ -Werten brauchen eine um so grössere Serienzahl r , um merkbar zu werden, je geringer $w(r)$ ist. Dies dürfte wohl einer der Hauptgründe dafür sein, dass man im Geltungsbereiche des »Gesetzes der kleinen Zahlen« relativ häufig eine Dispersion trifft, welche an der Hand des üblichen Forschungsverfahrens von der normalen nicht zu unterscheiden ist.

IV.

Besonders lehrreich kann sich das Heranziehen der höheren Potenzen der w_i -Zahlen bei den Dispersionsuntersuchungen dann gestalten, wenn die Versuchszahlen n_i bekannt sind.

Geht man hierbei von dem stochastischen Schema des dritten Kapitels aus, so lässt sich auf diesem Wege ein tieferer Einblick in die Verteilung der p_i -Werte gewinnen, indem, ausser dem Näherungswerte von $\frac{1}{r} \sum_{i=1}^r [p_i - p_{(r)}]^2$, auch Näherungswerte der höheren Momente der apriorischen p_i -Werte berechnet werden.

In ähnlicher Weise lässt sich gleichfalls jedes andere stochastische Schema behandeln.

Gelingt es hierbei, eine Anzahl von Gleichungen aufzustellen, welche die Zahl der das Schema charakterisierenden unbekannten Parameter übersteigt, so kommt man in die Lage, nicht nur die Werte der Constanten festzustellen, welche das empirische Material mit dem betr. Schema in Einklang setzen, sondern auch zu prüfen, ob sich das Material in das Schema widerspruchlos hineinzwingen lässt. Am einfachsten gestaltet sich die Prüfung eben für das Schema der normalen Dispersion, da man in diesem Falle bloss mit einer unbekannten Grösse — der allen w_i -Werten zu Grunde liegenden Wahrscheinlichkeit p — zu tun hat. Es genügt ja oft unter diesen Umständen die Berechnung von $\frac{1}{r-1} \sum_{i=1}^r n_i \left[\frac{1}{n_i} w_i - \frac{1}{s} \sum_{j=1}^r w_j \right]^2$, um die Unanwendbarkeit des Schemas nachzuweisen: ist nämlich die Differenz zwischen

$$\frac{1}{r-1} \sum_{i=1}^r n_i \left[\frac{1}{n_i} w_i - \frac{1}{s} \sum_{j=1}^r w_j \right]^2 \quad \text{und} \quad \frac{1}{s-1} \left[\sum_{i=1}^r w_i \right] \left[1 - \frac{1}{s} \sum_{i=1}^r w_i \right]$$

(bezw. die Differenz $Q^2 - 1$) grösser, als es ihrem mittleren Fehler entspricht, so lässt sich die Annahme nicht aufrechterhalten, dass die Voraussetzungen der normalen Dispersion tatsächlich erfüllt sind. Man erhält also eine definitive ver-

neinde Antwort auf die Frage, ob das Schema der normalen Dispersion zur Erklärung des Verhaltens der empirischen Zahlen herangezogen werden darf.

Fällt hingegen die Antwort bejahend aus, so kann sie, wie wir oben gesehen haben (S. 36 ff.), nicht ohne weiteres, als definitiv gelten: es bleibt stets die Möglichkeit bestehen, das Verhalten der empirischen Zahlen auf andere stochastische Schemata zurückzuführen. Verlangt man grössere Sicherheit, so bleibt kein anderer Weg offen, als eben die höheren Momente heranzuziehen, falls man sich im Rahmen der stochastischen Behandlung der gegebenen empirischen Zahlenreihe halten soll. Diesen Weg einzuschlagen, lohnt es sich freilich bloss, wenn die Zahl der Serien — r — hinreichend gross ist.

V.

Die stochastische Behandlung einer statistischen Zahlenreihe hat den Zweck, von den empirisch-»zufälligen« Zahlenwerten zu den ihnen zu Grunde liegenden »apriorischen Voraussetzungen« vorzudringen und dieselben, soweit es geht, zu entschleiern. Das nächste Ziel, das man sich hierbei vorsetzt, ist, die allgemeine Form des stochastischen Schemas, auf welches das Verhalten der empirischen Zahlen zurückgeführt werden kann, klarzulegen. An die Seite dieser »qualitativen« Aufgabe stellt sich dann die entsprechende »quantitative«, — die das betreffende Schema näher charakterisierenden Constanten mehr oder weniger genau zahlenmässig zu bestimmen.

Die ursprüngliche Fragestellung bei DORMOY und bei LEXIS, die in der Unterscheidung der normalen, der übernormalen und der unternormalen Dispersion gipfelt, betont mehr die »qualitative« Seite des Problems. Die »quantitative« Seite blieb zunächst so gut, wie unbeachtet. Dies ist ohne weiteres begreiflich. Im Falle der normalen Dispersion reduziert sich das »Quantitative« auf die Bestimmung der Grundwahrscheinlichkeit p , falls die Versuchszahlen, wie stets bei LEXIS, als bekannt gelten, — auf ein Problem mithin, das wegen seiner statistischen Alltäglichkeit, so zu sagen, im Vorbeigehen gelöst wird, ohne die Aufmerksamkeit auf seine

Eigenart zu ziehen. Und für die Fälle der übernormalen, sowie der unternormalen Dispersion verfügte man damals nicht über passende stochastische Schemata, deren zahlenmässiges Ausfüllen, als ein Problem für sich, erscheinen könnte.

In LEXIS' späterem Werk kommt die quantitative Forschungsrichtung zu ihren Rechten. Seine Theorie des Fehlerexcedenten geht bei der Behandlung der übernormalen Dispersion von einem bestimmten stochastischen Schema aus — dem Schema des dritten Kapitels dieser Abhandlung — und lässt in der »wesentlichen Schwankungscomponente« die wichtigste zahlenmässige Charakteristik des gewählten Schemas — den mittleren Fehler der Wahrscheinlichkeiten p_i — neben ihren Durchschnitt ($p_{(r)}$) in den Vordergrund treten. In gleicher Weise kann man von der allgemeineren Form desselben stochastischen Schemas ausgehen, die den Berechnungen des zweiten Kapitels dieser Abhandlung zu Grunde liegt, und

$m_{[1,r]}$ sowohl, wie $\frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2$, näherungsweise zu bestimmen suchen.

Man kann auf demselben Wege einen oder mehrere Schritte weiter gehen und höhere Momente der Grössen p_i (bezw. $m_i^{(i)}$) auf Grund der empirischen z_i -Werte zu berechnen suchen, um auf diese Weise eine tiefere Kenntnis der p_i (bezw. $m_i^{(i)}$)-Grössen zu erlangen. Die Aufgabe bleibt ihrem Wesen nach genau dieselbe, wie bei der Berechnung der LEXIS'schen »wesentlichen Schwankungscomponente«.

Auf gleichem Boden bleibt man auch, wenn man die Versuchszahlen, als unbekannt, betrachtet und unter Zugrundelegung eines bestimmten stochastischen Schemas — namentlich unter der Annahme der normalen Dispersion — Aufschlüsse über sie zu erhalten strebt. Diese Aufgabe ist gleichfalls ihrem Wesen nach stochastisch, und die Fragestellung, obgleich sie der PEARSON'schen Schule entstammt, fällt durchaus nicht aus dem Rahmen der Dispersionsuntersuchungen der LEXIS'schen Richtung.¹

¹ In dieser Beziehung kann ich mich an Prof. v. BORTKIEWICZ nicht anschliessen, der von einem »gegensätzlichen Verhältnisse« der PEARSON-WHITAKER'schen Methode (vgl. oben S. 42 ff.) zu dem von ihm selbst be-

Wird nun das gewählte stochastische Schema durch die Bestimmung der Näherungswerte der sie charakterisierenden Constanten ausgefüllt, so ist die stochastische Behandlung des statistischen Zahlenmaterials am Ende: ihr Leistungsvermögen ist hiermit erschöpft und ihre Aufgabe erfüllt. Bei der Wahl der stochastisch zu bestimmenden Parameter kann man sich durch verschiedene Gesichtspunkte leiten lassen. Geht man etwa von dem Schema des zweiten Kapitels dieser Abhandlung aus unter Voraussetzung, dass die n_i -Zahlen bekannt sind, so kann, als Aufgabe der stochastischen Behandlung der z_i -Werte, die Bestimmung der Näherungswerte der einzelnen $m_i^{(j)}$ -Grössen betrachtet werden. Andererseits kann man die Näherungswerte gewisser Funktionen derselben — des mittleren Fehlers, der höheren Momente — zu bestimmen suchen, — mit aus dem Grunde, dass die Spielräume der zufälligen Schwankungen für die einzelnen $m_i^{(j)}$ -Grössen zu gross sind. Man kann ferner in das stochastische Ausgangsschema die Annahme einer bestimmten Form des Verteilungsgesetzes der $m_i^{(j)}$ -Werte aufnehmen und die Bestimmung der Näherungswerte der Constanten dieses Gesetzes als die eigentliche Aufgabe der stochastischen Behandlung des statistischen Zahlenmaterials betrachten. Das supponierte Gesetz kann hierbei sowohl die Form eines »Verteilungsgesetzes« im Sinne der Häufigkeitscurve, d. i. eines analytischen Ausdrucks für den Zusammenhang zwischen der Grösse und der relativen Anzahl der $m_i^{(j)}$ -Werte, haben, wie auch die Form eines analytischen Ausdrucks für den Zusammenhang zwischen der Grösse von $m_i^{(j)}$ und der Ordnungsnummer i annehmen, wie in den meisten Fällen der Interpolation nach der Methode der kleinsten Quadrate.

Bei einer solchen Fragestellung greifen stochastische und interpolatorische Aufgaben unmittelbar ineinander, und das »Gesetz«, dem die $m_i^{(j)}$ -Grössen folgen, erscheint, als fertiges Product der stochastischen Bearbeitung des Zahlenmaterials. Dies ist das gemeinsame Kennzeichen der Interpolation nach der Methode der kleinsten Quadrate, des interpolatorischen

vorzugten Verfahren (Realismus und Formalismus, S. 227), ja sogar von einem »neuen geradezu revolutionären Prinzip der schematischen Bestimmung der Grundzahl« (ibid., S. 239) spricht.

Verfahrens von CAUCHY, der PEARSON'schen Methoden der Frequenzcurvenberechnung u. s. w.

Die stochastische Behandlung des Zahlenmaterials lässt sich aber wohl von interpolatorischen Nebenabsichten ganz frei halten, indem man keine Annahmen über die Form des »Gesetzes« der $m_1^{(i)}$ -Werte in das stochastische Ausgangsschema aufnimmt. Dann trennen sich die stochastischen und die interpolatorischen Aufgaben scharf von einander. Als Endzweck der stochastischen Arbeit erscheint dann eben die Bestimmung der Näherungswerte von $m_{[1,r]}$ von

$\frac{1}{r} \sum_{i=1}^r [m_1^{(i)} - m_{[1,r]}]^2$ u. s. w., und, als eine gesonderte Aufgabe

für sich, lässt sich eventuell die Bestimmung des Verteilungsgesetzes von $m_1^{(i)}$ auf Grundlage der stochastisch bestimmten Näherungswerte einer Reihe von Momenten derselben betrachten. Dieses »problème des moments«, wie es im Anschluss an STIELTJES¹ genannt werden mag, hat an sich mit der Stochastik nichts gemein: die Wege, die hier zu Lösungen führen, halten sich von der Wahrscheinlichkeitsrechnung fern. Aber die Ergebnisse, die auf diesem eigenartigen Gebiete einer Interpolationsrechnung, welche auf willkürliche Voraussetzungen verzichtet, erzielt werden, sind von allergrösstem Werte für den Ausbau der stochastischen Theorie der Statistik: es sei hier bloss an die grundlegenden TCHEBYCHEFF'schen Ungleichheiten² erinnert.

¹ Cf. STIELTJES, Recherches sur les fractions continues (Annales de la faculté de Toulouse, 1894).

² Cf. TCHEBYCHEFF, Oeuvres, t. II, p. 478 (russische Auflage): Wenn eine Funktion $f(x)$ stets positiv bleibt und

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx &= 1, \\ \int_{-\infty}^{+\infty} x^{2i-1} f(x) dx &= 0 \text{ für } i = 1, 2, \dots, m, \\ \int_{-\infty}^{+\infty} x^{2i} f(x) dx &= \frac{1 \cdot 3 \cdot 5 \dots (2i-1)}{q^{2i}} \text{ für } i = 1, 2, \dots, (m-1), \end{aligned}$$

Das Fernhalten der Interpolationsrechnungen ist ein Kennzeichen der stochastischen Arbeiten der sozialwissenschaftlich orientierten LEXIS'schen Richtung. Ein lebhaftes Interesse für die Auffindung der »Gesetze« kennzeichnet die an den Naturwissenschaften orientierte PEARSON'sche Schule. In der Beurteilung des Erkenntniswertes der interpolatorisch zu erzielenden Ergebnisse gehen die beiden Richtungen zweifellos ziemlich weit auseinander. Soweit aber die Arbeiten der PEARSON'schen Schule den stochastischen Boden nicht verlassen, darf m. E. von einem tieferen Gegensatz zur LEXIS'schen Richtung nicht gesprochen werden. Unterschiede — und zwar zum Teil schwerwiegende — sind gewiss vorhanden. Die Vertreter der englischen Schule haben von Anbeginn in ihren stochastischen Untersuchungen das Quantitative in den Vordergrund geschoben, ohne sich um die qualitativen Voraussetzungen der auszuführenden Messungen in erforderlicher Masse zu kümmern. Viel Unheil ist ferner durch die Abneigung der englischen Forscher gegen die Begriffe der mathematischen Wahrscheinlichkeit und der mathematischen Erwartung gestiftet worden: der Verzicht auf den Gebrauch dieser Grundbegriffe hat der Klarheit der stochastischen Fragestellungen ungemein geschadet, — ja gelegentlich die Lösungsversuche auf Irrwege gelenkt. Legt man jedoch das, dem kontinentalen Auge nicht zusagende Gewand ab und holt, wo dies erforderlich ist, das Versäumte nach, so zeigt sich deutlich, dass PEARSON und LEXIS zur Lösung wesens-

so ist $\int_{-\infty}^v f(x) dx$ in den Grenzen

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{q}{\sqrt{2}}} e^{-x^2} dx \pm \frac{3 \sqrt{3} [m^2 - 2m + 3]^{\frac{3}{2}} [q^2 v^2 + 1]^3}{2 [m - 3]^3 \sqrt{m - 1}}$$

enthalten. Mit wachsendem m streben die beiden Grenzen demselben

Grenzwerte $\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{q}{\sqrt{2}}} e^{-x^2} dx$ zu, und die Verteilung unterscheidet sich somit

bei $m = \infty$ nicht von der GAUSS'schen.

verwandter stochastischer Aufgaben der Form nach oft verschiedene, aber im Grunde doch wesensverwandte Verfahren einschlagen. Als eine der wichtigsten Aufgaben der aufblühenden stochastischen Theorie der Statistik erscheint gegenwärtig eben, das von beiden Förderern derselben Erzielte zu einer die Gegensätze aufhebenden Synthese zu verarbeiten. Nicht »LEXIS gegen PEARSON«, sondern »PEARSON durch LEXIS geläutert, LEXIS durch PEARSON bereichert« sollte gegenwärtig die Parole derer lauten, die, von der geistlosen Empirie der nachqueteletistischen Statistik unbefriedigt, sich nach einer rationellen Theorie der Statistik sehnen.¹

¹ Die obigen Betrachtungen sind im Anschluss an den scharfen Angriff formuliert worden, den Prof. v. BORTKIEWICZ gegen die PEARSON'sche Richtung in der Abhandlung »Realismus und Formalismus in der math. Statistik« gerichtet hat. Was »den tiefen Gegensatz der beiden (i. e. der realistischen und der formalistischen) Richtungen« anbelangt, so will ich ihn keineswegs in Abrede stellen, falls unter Formalismus »diejenige Richtung in der math. Statistik, welche in der Auffindung descriptiver Formeln die Hauptaufgabe der Wissenschaft erblickt« (Real. u. Form., S. 245), verstanden werden soll, und nichts liegt mir ferner, als für einen derartigen »Formalismus« eintreten zu wollen. Ich glaube aber, dass der Boden dieser Definition verlassen wird, wenn z. B. das »schematische« PEARSON-WHITAKER'sche Verfahren zur Bestimmung der Versuchszahl ohne weiteres, als formalistisch, dem von BORTKIEWICZ bevorzugten »realistischen« Verfahren gegenüber gestellt wird (S. 252). BORTKIEWICZ tut L. WHITAKER Unrecht, wenn er ihre Untersuchung, als durch und durch »formalistisch« im obigen Sinne, brandmarkt (S. 245): die Abhandlung von L. WHITAKER enthält einen kräftigen »realistischen« Kern und stellt einen bedeutsamen Beitrag zur Dispersionstheorie dar. Was dann K. PEARSON selbst anbelangt, so wäre es vollends ungerecht, ihn zu den »Formalisten« zu rechnen aus dem Grunde, dass in seinem reichhaltigen »oeuvre« auch wertvolle Beiträge zur Lösung der »formalistischen« Interpolationsaufgaben enthalten sind.