Wiss. Z. Univ. Halle XXXII'83 M, H. 5, S. 87—89

Über eine Partition der nat. Zahlen und ihre Anwendung beim U-Test

Andreas Löffler

Der U-Test von Mann und Whitney (ein parameterfreies Prüfverfahren zum Vergleich zweier unabhängiger Stichproben mit den Umfängen n_1 und n_2 ($n_2 \geqq n_1$)) ist wegen seiner einfachen Handhabung ein häufig verwendeter Test. Für kleine Stichprobenumfänge ($n_2 < 20$) mußten aber bisher die kritischen Werte der Prüfgröße speziellen Tabellen entnommen werden [1]. Vor allem bei Rechenprogrammen für Kleinstrechner mit geringer Speicherkapazität war dies recht aufwendig. Das Problem bestand nun darin, ein rationelleres Verfahren für die Berechnung dieser kritischen Werte zu finden, d. h. die Anzahl der Möglichkeiten der Anordnung der einzelnen Elemente aus den Stichproben, zu denen jeweils der gleiche Wert der Prüfgröße gehört, zu berechnen. Man gelangt hierbei zu folgender zahlentheoretischer Aufgabenstellung: Auf wie viele verschiedene Weisen läßt sich eine nat. Zahl a als Summe von n_1 nat. Zahlen, welche höchstens gleich n_2 sind, darstellen (Partition mit Einschränkungen)?
Jede Zerlegung von a der Form

$$a = a_1 + a_2 + \ldots + a_{n_1}; \quad 0 \leqq a_1 \leqq \ldots \leqq a_{n_1} \leqq n_2, \quad a_\nu \in N \tag{1}$$

wird Zerfällung von a genannt. Die Anzahl der Zerfällungen (1) von a wird mit $p_{n_1,n_2}(a)$ bezeichnet. Dann gilt nach [2]

$$p_{n_1,n_2}(a) = p_{n_2,n_1}(a)\,. \tag{2}$$

Zur Berechnung von $p_{n_1,n_2}(a)$ kann man sich also auf $n_2 \geqq n_1$ beschränken. Durch eine Modifizierung des Beweisweges von [2] erhält man leicht

$$p_{n_1,n_2}(a) = p_{n_2,n_1}(n_1 n_2 - a)\,.$$

Eine einfache Rekursionsformel für die Berechnung der Partition gibt [3] an:

$$p_{n_1,n_2}(a) = p_{n_1-1,n_2}(a - n_2) + p_{n_1,n_2-1}(a)\,. \tag{3}$$

Mit Hilfe von (2) und (3) soll jetzt eine erzeugende Funktion $F_{n_1,n_2}(z)$ ($|z| < 1$) für die Partition $p_{n_1,n_2}(a)$ gefunden werden. Das ist diejenige reellwertige Funktion, die eine Reihendarstellung besitzt, bei der die $p_{n_1,n_2}(a)$ gerade die Koeffizienten der z^a darstellen. Da $p_{n_1,n_2}(a) = 0$, wenn $a > n_1 n_2$ ist, ist also

$$F_{n_1,n_2}(z) = \sum_{a=0}^{n_1 n_2} p_{n_1,n_2}(a)\, z^a\,. \tag{4}$$

Wegen der Darstellung der Zerfällungen nach (1) gilt weiter

$$F_{n_1,n_2}(z) = \sum_{a_1=0}^{n_2} \sum_{a_2=a_1}^{n_2} \cdots \sum_{a_{n_1}=a_{n_1-1}}^{n_2} z^{a_1+\ldots a_{n_1}} \tag{4}$$

Es gilt folgender

Satz 1:

$$F_{n_1,n_2}(z)=\prod_{\nu=1}^{n_1}\left(\frac{1-z^{n_2+\nu}}{1-z^\nu}\right). \tag{5}$$

Beweis: Der Beweis wird geführt mit vollständiger Induktion nach n_1 und n_2. Wegen (2) ist im Induktionsanfang nur die Gültigkeit für $p_{1,n_2}(a)$ zu zeigen:

$$\sum_{a=0}^{n_2} p_{1,n_2}(a)\, z^a=\sum_{a=0}^{n_2} z^a=\frac{1-z^{n_2+1}}{1-z}.$$

Der Induktionsschritt wird mit (3) geführt:

$$\begin{aligned}\sum_{a=0}^{n_1n_2} p_{n_1,n_2}(a)\, z^a &= z^{n_2}\sum_{a=0}^{n_1n_2} p_{n_1-1,n_2}(a-n_2)\, z^{a-n_2}+\sum_{a=0}^{n_1n_2} p_{n_1,n_2-1}(a)\, z^a\\ &= z^{n_2}\prod_{\nu=1}^{n_1-1}\left(\frac{1-z^{n_2+\nu}}{1-z^\nu}\right)+\prod_{\nu=1}^{n_1}\left(\frac{1-z^{n_2-1+\nu}}{1-z^\nu}\right)\\ &= \prod_{\nu=1}^{n_1}\left(\frac{1-z^{n_2+\nu}}{1-z^\nu}\right) \quad \text{w. z. b. w.}\end{aligned}$$

Jetzt soll eine Rekursionsformel hergeleitet werden, bei der im Gegensatz zu (3) nur die Partitionen $p_{n_1,n_2}(i)$ benötigt werden. Definiert man folgende Teilerfunktion

$$\sigma(n;n_1,n_2)=\sum_{d|n}\varepsilon_d\, d \quad \text{mit} \quad \varepsilon_d=\begin{cases}1, \text{wenn } 1\leqq d\leqq n_1\\ 0, \text{sonst}\\ -1, \text{wenn } n_2+1\leqq d\leqq n_2+n_1,\end{cases} \tag{6}$$

dann gilt folgender

Satz 2:

$$a p_{n_1,n_2}(a)=\sum_{i=0}^{a-1} p_{n_1,n_2}(i)\,\sigma(a-i;n_1,n_2). \tag{7}$$

Beweis: Dazu geht man aus von der Identität

$$\begin{aligned}\frac{F'_{n_1,n_2}(z)}{F_{n_1,n_2|z|}} &= \frac{d}{dz}(\ln F_{n_1,n_2}(z))\\ &= \sum_{\nu=1}^{n_2}\left(\frac{\nu z^{\nu-1}}{1-z^\nu}-\frac{(n_2+\nu)\, z^{n_2+\nu-1}}{1-z^{n_2+\nu}}\right)\\ &= \sum_{\nu=1}^{n_1}\sum_{m=1}^{\infty}\left(\nu z^{m\nu-1}-(n_2+\nu)\, z^{(n_2+\nu)m-1}\right)\\ &= \sum_{n=1}^{\infty}\left(\left(\sum_{\substack{\nu m=n\\ 1\leqq\nu\leqq n_1}}\nu\right)z^{n-1}-\left(\sum_{\substack{\nu m=n\\ n_2+1\leqq\nu\leqq n_2+n_1}}\nu\right)z^{n-1}\right)\\ &= \sum_{n=1}^{\infty}\sigma(n;n_1,n_2)\, z^{n-1}.\end{aligned}$$

Daraus ergibt sich

$$\sum_{a=1}^{n_1n_2} a p_{n_1,n_2}(a)\, z^{a-1}=\sum_{a=0}^{n_1n_2}\sum_{n=1}^{\infty} p_{n_1,n_2}(a)\,\sigma(n;n_1,n_2)\, z^{a+n-1}$$

und mit Koeffizientenvergleich die Behauptung des Satzes. w. z. b. w.

Mit (7) ist eine schnelle und nicht so speicheraufwendige Berechnung der kritischen Werte möglich.

Anmerkung: [2] gibt für andere Partitionen erzeugende Funktionen an. Man sieht sofort, daß

$$F_{\mathfrak{A}}(z) = (1 + z^{a_0} + z^{2a_0} + \dots)(1 + z^{a_1} + z^{2a_1} + \dots)\dots$$

$$= \prod_{a_\nu \in \mathfrak{A}} (1 - z^{a_\nu})^{-1}$$ die Partition mit Summanden aus der Menge $\mathfrak{A} = \{a_0, a_1, \dots\}$, $a_\nu \in N$, erzeugt,

$$F_k(z) = \prod_{\nu=1}^{k} (1 - z^\nu)^{-1}$$ die Partition mit k als höchsten Summanden oder höchstens k Summanden (nach (2)) erzeugt,

$$F(z) = \prod_{\nu=1}^{\infty} (1 - z^\nu)^{-1}$$ die Partition ohne Einschränkung erzeugt,

$$F_v(z) = \prod_{\nu=1}^{\infty} (1 + z^\nu)$$ die Partition mit paarweise verschiedenen Summanden erzeugt.

Rekursionsformeln erhält man hier durch rekursive Darstellung der erzeugenden Funktion ($F_k(z) = (1 - z^k)F_{k-1}(z)$) und Koeffizientenvergleich sowie auf analogem Wege zu Satz 2.

LITERATUR

[1] Müller, P.-H., u. a.: Tafeln der math. Statistik. Leipzig 1973.

[2] Ostmann, H.-H.: Additive Zahlentheorie, 1. Teil (allg. Untersuchungen). Berlin-Göttingen-Heidelberg 1956, S. 32ff.

[3] Fisz, M.: Wahrscheinlichkeitsrechnung und math. Statistik, Berlin 1978, S. 524f.

Manuskripteingang: 3. 8. 1982

Verfasser:

Andreas Löffler, Spezialklasse an der Sektion Mathematik der Martin-Luther-Universität Halle-Wittenberg