Nachtrag zu Abschnitt 2

Andreas Eberle
Institut für angewandte Mathematik

Oktober 2008

Nachtrag/Exkurs zu Abschnitt 2

- Angenommen, wir wollen feststellen, ob ein Merkmal X jede seiner möglichen Ausprägungen $a_1, a_2, \ldots, a_r$ mit derselben Wahrscheinlichkeit annimmt. Beipielsweise wollen wir testen, ob ein Würfel fair oder gezinkt ist.
- Sind alle Ausprägungen gleich wahrscheinlich, dann beträgt die Wahrscheinlichkeit jeder der r möglichen Ausprägungen $a_1, a_2, \ldots, a_r$ gerade
$$p = \frac{1}{r}$$
- In einer genügend großen ("repräsentativen") Stichprobe sollte dann auch die relative Häufigkeit der Ausprägung a_k ungefähr gleich diesem Wert sein ("Gesetz der großen Zahl", s.u.)
$$h_k = h(a_k) \approx p \qquad \text{bzw.} \qquad \frac{h_k}{p} \approx 1 \qquad (*)$$
für $k = 1, 2, \ldots, r$.
- Dies gilt aber in der Regel nicht exakt, da die relativen Häugigkeiten zufälligen Fluktuationen unterliegen.

Nachtrag zu Abschnitt 2

Chiquadrat-Statistik

Fakt: *Haben alle a_k dieselbe Wahrscheinlichkeit $p = \frac{1}{r}$, dann sollte $\frac{h_k}{p} \approx 1$ gelten.*

- Also liegt es nahe, die mittlere quadratische Abweichung

$$\phi^2 = \frac{1}{r} \sum_{k=1}^{r} \left(\frac{h_k}{p} - 1 \right)^2 = \sum_{k=1}^{r} \frac{(h_k - p)^2}{p} = \frac{1}{n} \sum_{k=1}^{r} \frac{(n_k - n \cdot p)^2}{n \cdot p}$$

  als Maß für die Abweichung von der Gleichverteilung zu betrachten.

Definition

Die χ^2-**Statistik** ist

$$\chi^2 = \sum_{k=1}^{r} \frac{(n_k - n \cdot p)^2}{n \cdot p}$$

Hierbei ist $n \cdot p$ die bei Wahrscheinlichkeit p im Schnitt zu erwartende Häufigkeit einer Merkmalsausprägung.

Beispiel. (Zufall oder nicht ?)

- *Ein Mitspieler erzielt bei 30 mal Würfeln folgende Augenzahlen:*

1	2	3	4	5	6
4	3	3	5	4	11

  Können Sie schließen, daß manipuliert wird ?

- *Berechnen von* χ^2 *ergibt*

$$\begin{aligned}\chi^2 &= \frac{(11-5)^2}{5} + \frac{(4-5)^2}{5} + \cdots + \frac{(4-5)^2}{5} \\ &= 7.2 + 0.2 + \cdots + 0.2 = 9.2\end{aligned}$$

- *Man kann zeigen, daß bei einem fairen Würfel der Wert dieser Chiquadrat-Statistik mit ca. 10 % Wahrscheinlichkeit oberhalb von* 9.2 *liegt, s.u.*
- *Die Häufigkeitsverteilung könnte also durchaus durch zufällige Fluktuationen entstanden sein - wir können nicht auf Manipulation schließen.*

- Sie bleiben trotzdem mißtrauisch, und notieren die Häufigkeiten bei 600 Würfen:

1	2	3	4	5	6
94	81	107	87	93	138

- Berechnen von χ^2ergibt nun

$$\begin{aligned}\chi^2 &= \frac{(138-100)^2}{100}+\frac{(93-100)^2}{100}+\cdots+\frac{(94-100)^2}{100} \\ &= 14.44+0.49+1.69+0.49+3.61+0.36 \;= 21.08\end{aligned}$$

- Die Wahrscheinlichkeit, daß die Chiquadratstatistik bei Gleichverteilung einen Wert oberhalb von 21 annimmt ist kleiner als $\frac{1}{1000}$. Sie vermuten, daß hier etwas nicht stimmt.
- Um dies zu überprüfen könnten Sie die Chiquadratstatistik für 600 weitere Würfe berechnen - die Wahrscheinlichkeit, daß sich zweimal hintereinander ein so hoher Wert ergibt, ist kleiner als ein Millionstel.