Die Hypergeometrische Verteilung

Timm Grams, Fulda, 31. Juli 2010 (rev. 02.08.10)

Formel für die Wahrscheinlichkeiten

Die Grundmenge der Größe N möge M Merkmalsträger enthalten. Es ist $M \leq N$. Aus dieser Grundmenge werden n Elemente rein zufällig ausgewählt. Die Wahrscheinlichkeit p_k, dass sich in dieser Auswahl genau k Merkmalsträger befinden, ist gleich

$$p_k = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}.$$

Dass es sich hier tatsächlich um eine Verteilung handelt, wird dadurch nachgewiesen, dass die Summe der p_k gleich 1 ist. Das ist genau dann der Fall, wenn die Formel

$$\sum_{k=0}^{n} \binom{M}{k} \cdot \binom{N-M}{n-k} = \binom{N}{n}$$

gilt. Ich beweise sie mittels vollständiger Induktion über N. (Ich reite auf der Sache mehr als nötig herum, weil wir die Formel in den folgenden Herleitungen wiederholt benötigen werden.)

Induktionsanfang: Weil 0 über 0 gleich 1 ist, gilt die Formel für $N = 0$.

Induktionsschluss: Wie man leicht einsieht, gilt die Formel für M=0 grundsätzlich. Wir können für den Induktionsschluss also davon ausgehen, dass M eine natürliche Zahl ist, dass also 0<M gilt. Nun nehmen wir an, dass die Formel für N-1 bereits bewiesen worden ist. Daraus folgt

$$\sum_{k=0}^{n} \binom{M}{k} \cdot \binom{N-M}{n-k} = \binom{N-M}{n} + \sum_{k=1}^{n} \binom{M}{k} \cdot \binom{N-M}{n-k}$$

$$= \binom{N-M}{n} + \sum_{k=1}^{n} \left(\binom{M-1}{k-1} + \binom{M-1}{k} \right) \cdot \binom{N-M}{n-k}$$

$$= \sum_{k=1}^{n} \binom{M-1}{k-1} \cdot \binom{N-M}{n-k} + \sum_{k=0}^{n} \binom{M-1}{k} \cdot \binom{N-M}{n-k}$$

$$= \sum_{k=1}^{n} \binom{M-1}{k-1} \cdot \binom{N-M}{n-k} + \sum_{k=0}^{n} \binom{M-1}{k} \cdot \binom{N-M}{n-k}$$

$$= \sum_{k=1}^{n} \binom{M-1}{k-1} \cdot \binom{N-1-(M-1)}{n-1-(k-1)} + \sum_{k=0}^{n} \binom{M-1}{k} \cdot \binom{N-1-(M-1)}{n-k}$$

$$= \sum_{k=0}^{n-1} \binom{M-1}{k} \cdot \binom{N-1-(M-1)}{n-1-k} + \sum_{k=0}^{n} \binom{M-1}{k} \cdot \binom{N-1-(M-1)}{n-k}$$

$$= \binom{N-1}{n-1} + \binom{N-1}{n} = \binom{N}{n},$$

also die Gültigkeit der Formel für N.

Der Erwartungswert

Der Erwartungswert μ ist in diesem Fall gegeben durch $\mu = \sum_{k=0}^{n} k \cdot p_k$. Wegen

$$\mu = \sum_{k=0}^{n} x_k p_k = \sum_{k=0}^{n} k \cdot \binom{M}{k} \cdot \binom{N-M}{n-k} \Big/ \binom{N}{n} = \sum_{k=1}^{n} k \cdot \binom{M}{k} \cdot \binom{N-M}{n-k} \Big/ \binom{N}{n}$$

$$= \sum_{k=1}^{n} M \cdot \binom{M-1}{k-1} \cdot \binom{N-M}{n-k} \Big/ \binom{N}{n}$$

$$= \sum_{k=1}^{n} M \cdot \binom{M-1}{k-1} \cdot \binom{N-1-(M-1)}{n-1-(k-1)} \Big/ \binom{N}{n}$$

$$= \sum_{k=0}^{n-1} M \cdot \binom{M-1}{k} \cdot \binom{N-1-(M-1)}{n-1-k} \Big/ \binom{N}{n}$$

$$= M \cdot \sum_{k=0}^{n-1} \binom{M-1}{k} \cdot \binom{N-1-(M-1)}{n-1-k} \Big/ \binom{N}{n}$$

$$= M \cdot \binom{N-1}{n-1} \Big/ \binom{N}{n} = n \frac{M}{N}.$$

ist also

$$\boxed{\mu = n \frac{M}{N}}$$

Die Streuung

Die Streuung σ^2 ist in diesem Fall gegeben durch $\sigma^2 = \sum_{k=0}^{n} (k-\mu)^2 p_k = \left(\sum_{k=1}^{n} k^2 p_k \right) - \mu^2$. Es ist

$$\sum_{k=1}^{n} k^2 p_k = \sum_{k=1}^{n} \frac{k^2 \cdot \binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

$$= \sum_{k=1}^{n} \frac{k(k-1) \cdot \binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} + \sum_{k=1}^{n} \frac{k \cdot \binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

$$= \sum_{k=2}^{n} \frac{k(k-1) \cdot \binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} + n \frac{M}{N}$$

$$= \sum_{k=2}^{n} \frac{M(M-1) \cdot \binom{M-2}{k-2} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} + n \frac{M}{N}$$

$$= \sum_{k=2}^{n} \frac{M(M-1) \cdot \binom{M-2}{k-2} \cdot \binom{N-2-(M-2)}{n-2-(k-2)}}{\binom{N}{n}} + n\frac{M}{N}$$

$$= M(M-1) \cdot \sum_{k=0}^{n-2} \frac{\binom{M-2}{k} \cdot \binom{N-2-(M-2)}{n-2-k}}{\binom{N}{n}} + n\frac{M}{N}$$

$$= \frac{M(M-1)}{\binom{N}{n}} \cdot \sum_{k=0}^{n-2} \binom{M-2}{k} \cdot \binom{N-2-(M-2)}{n-2-k} + n\frac{M}{N}$$

$$= \frac{M(M-1)}{\binom{N}{n}} \cdot \binom{N-2}{n-2} + n\frac{M}{N}$$

$$= \frac{M(M-1)n(n-1)}{N(N-1)} + n\frac{M}{N}.$$

Für die Streuung ergibt sich damit

$$\sigma^2 = \frac{M(M-1)n(n-1)}{N(N-1)} + n\frac{M}{N} - \left(n\frac{M}{N}\right)^2$$

$$= \frac{M(M-1)n(n-1)}{N(N-1)} + n\frac{M}{N} - \left(n\frac{M}{N}\right)^2$$

$$= \frac{MN(M-1)n(n-1) + MN(N-1)n - M^2(N-1)n^2}{N^2(N-1)}$$

$$= \frac{MN(M-1)n(n-1) + MN(N-1)n - M^2(N-1)n^2}{N^2(N-1)}$$

$$= \frac{M^2Nn^2 - M^2Nn - MNn^2 + MNn + MN^2n - MNn - M^2Nn^2 + M^2n^2}{N^2(N-1)}$$

$$= \frac{-M^2Nn - MNn^2 + MN^2n + M^2n^2}{N^2(N-1)}$$

$$= \frac{MN(N-M)n - M(N-M)n^2}{N^2(N-1)}$$

$$= \frac{nM(N-M)(N-n)}{N^2(N-1)}.$$

In Kürze:

$$\boxed{\sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)}}$$