Varianz der hypergeometrischen Verteilung

Die Zufallsvariable X sei hypergeometrisch verteilt mit Parametern N, M und n. Dann gilt $X(\Omega) = \{0, 1, 2, \dots, n\}$ und wir haben bereits gezeigt, dass $\mathbb{E}(X) = n\frac{M}{N}$. Wir zeigen

$$\mathbb{V}(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}.$$

Beweis. Es ist

$$\begin{aligned}
\binom{N}{n} \cdot \mathbb{V}(X) &= \sum_{k=0}^{n}\left(k - n\frac{M}{N}\right)^2 \binom{M}{k}\binom{N-M}{n-k} \\
&= \sum_{k=0}^{n}\left(k^2 - 2n\frac{M}{N}k + n^2\frac{M^2}{N^2}\right)\binom{M}{k}\binom{N-M}{n-k} \\
&= \sum_{k=0}^{n}\left(k(k-1) + \left(1 - 2n\frac{M}{N}\right)k + n^2\frac{M^2}{N^2}\right)\binom{M}{k}\binom{N-M}{n-k} \\
&= \sum_{k=0}^{n} k(k-1)\binom{M}{k}\binom{N-M}{n-k} + \left(1 - 2n\frac{M}{N}\right)\sum_{k=0}^{n} k\binom{M}{k}\binom{N-M}{n-k} \\
&\qquad + n^2\frac{M^2}{N^2}\sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k}.
\end{aligned}$$

Die zweite der drei auftetenden Summen kann man aus der Berechnung des Erwartungswerts übernehmen:

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

Auf die dritte Summe kann man das Additionstheorem für Binomialkoeffizienten anwenden:

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

Setzt man beides ein, so erhält man

$$\binom{N}{n} \cdot \mathbb{V}(X) = \sum_{k=0}^{n} k(k-1)\binom{M}{k}\binom{N-M}{n-k} + \left(1 - 2n\frac{M}{N}\right)M\binom{N-1}{n-1} + n^2\frac{M^2}{N^2}\binom{N}{n}.$$

Die erste Summe kann man folgendermaßen berechnen:

$$\begin{aligned}
&\sum_{k=0}^{n} k(k-1)\binom{M}{k}\binom{N-M}{n-k} = \sum_{k=2}^{n} k(k-1)\binom{M}{k}\binom{N-M}{n-k} \\
&= \sum_{k=2}^{n} M(M-1)\binom{M-2}{k-2}\binom{N-M}{n-k} = M(M-1)\sum_{k=2}^{n}\binom{M-2}{k-2}\binom{N-M}{n-k} \\
&= M(M-1)\sum_{k=0}^{n-2}\binom{M-2}{k}\binom{(N-2)-(M-2)}{n-2-k} = M(M-1)\binom{N-2}{n-2}
\end{aligned}$$

Dabei wurde im letzten Schritt wieder das Additionstheorem für Binomialkoeffizienten verwendet. Setzt man das oben ein, so erhält man

$$\binom{N}{n} \cdot \mathbb{V}(X) = M(M-1)\binom{N-2}{n-2} + \left(1 - 2n\frac{M}{N}\right)M\binom{N-1}{n-1} + n^2\frac{M^2}{N^2}\binom{N}{n}.$$

Nun ist

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N(N-1)}{n(n-1)}\frac{(N-2)!}{(n-2)!((N-2)-(n-2))!} = \frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}$$

und daher

$$\begin{aligned}\frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}\mathbb{V}(X) = M(M-1)\binom{N-2}{n-2} + \left(1 - 2n\frac{M}{N}\right)M\frac{N-1}{n-1}\binom{N-2}{n-2} \\ + n^2\frac{M^2}{N^2}\frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}.\end{aligned}$$

Nach Kürzen durch den Binomialkoeffizienten $\binom{N-2}{n-2}$ ergibt sich daraus

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

und daher

$$\begin{aligned}\mathbb{V}(X) &= M(M-1)\frac{n(n-1)}{N(N-1)} + \left(1 - 2n\frac{M}{N}\right)M\frac{n}{N} + n^2\frac{M^2}{N^2} \\ &= n\frac{M}{N}\left(\frac{(M-1)(n-1)}{N-1} + \frac{N-2nM}{N} + \frac{nM}{N}\right) \\ &= n\frac{M}{N}\left(\frac{(M-1)(n-1)}{N-1} + \frac{N-nM}{N}\right) \\ &= n\frac{M}{N}\frac{N(M-1)(n-1) + (N-nM)(N-1)}{N(N-1)} \\ &= n\frac{M}{N}\frac{NMn - NM - Nn + N + N^2 - NMn - N + nM}{N(N-1)} \\ &= n\frac{M}{N}\frac{N^2 - NM - Nn + nM}{N(N-1)} \\ &= n\frac{M}{N}\frac{(N-M)(N-n)}{N(N-1)} \\ &= n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}.\end{aligned}$$

Die obige Rechnung kann man für $n = 1$ so nicht durchführen. Wir rechnen diesen Fall darum extra:

$$\begin{aligned}N \cdot \mathbb{V}(X) &= \binom{N}{1} \cdot \mathbb{V}(X) = \sum_{k=0}^{1}\left(k - \frac{M}{N}\right)^2\binom{M}{k}\binom{N-M}{1-k} \\ &= \frac{M^2}{N^2}(N-M) + \left(1 - \frac{M}{N}\right)^2 M = \frac{M^2(N-M) + (N-M)^2M}{N^2} \\ &= \frac{M^2N - M^3 + N^2M - 2NM^2 + M^3}{N^2} = \frac{N^2M - NM^2}{N^2} = \frac{NM - M^2}{N}\end{aligned}$$

Daraus folgt

$$\begin{aligned}\mathbb{V}(X) &= \frac{NM - M^2}{N^2} = \frac{M(N-M)}{N^2} = \frac{M}{N}\left(1 - \frac{M}{N}\right) \\ &= 1 \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-1}{N-1} = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1},\end{aligned}$$

d.h. die obige Formel ist auch für $n = 1$ korrekt.