4. Spezielle Verteilungen

<u>Lernziele</u>:

- Sie kennen verschiedene spezielle Verteilungen, ihre charakteristischen Eigenschaften und die entsprechenden R-Funktionen.
- Sie finden für ein Anwendungsproblem ein passendes stochastisches Modell, d.h. die spezielle Verteilung, deren Eigenschaften das Problem am besten beschreiben.

4.1 Bernoulli- und Binomialverteilung

4.1.1 Bernoulli-Verteilung

<u>Situation</u>: Experiment mit den Ausgängen "Erfolg" oder "Misserfolg"

Man beobachtet, ob ein bestimmtes Ereignis A eintritt oder nicht.
Die zugehörige ZV X ist ein Indikator für den Eintritt von A.

$$X = \begin{cases} 1, & \text{falls A eintritt (Erfolg)} \\ 0, & \text{falls A nicht eintritt (Misserfolg)} \end{cases}$$

<u>Verteilung</u>: $X \overset{\text{hat Verteilung}}{\sim} B_{1,p}$ mit p: Erfolgswahrscheinlichkeit

<u>Wahrscheinlichkeitsfkt.</u>:

$$P_{B_{1,p}}(x) = p^x \cdot (1-p)^{1-x} = \begin{cases} p, & \text{falls } x=1 \\ 1-p, & \text{falls } x=0 \\ 0, & \text{sonst} \end{cases}$$

Erwartungswert und Varianz:

$$E[X] = 1 \cdot p + 0 \cdot (1-p) = p$$

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$$

4.1.2 Binomialverteilung (diskrete ZV)

Situation: n unabhängige und identische Bernoulli-Experimente, d.h. n-maliges Ziehen mit Zurücklegen aus $\{0,1\}$

Man beobachtet die Anzahl der Erfolge, d.h. die ZV X mit Werten $0,1,\dots,n$.

Verteilung: $X \sim B_{n,p}$, n: Anzahl der Wiederholungen
p: Erfolgswahrscheinlichkeit

Wahrscheinlichkeitsfkt.:

$$P_{B_{n,p}}(x) = \underbrace{\binom{n}{x}}_{\text{Anzahl der Möglichkeiten mit } x \text{ Erfolgen}} p^x \cdot (1-p)^{n-x} \quad \text{für } x \in \{0,1,\dots,n\}$$

Erwartungswert und Varianz: n-malige Durchführung von unabh. Bernoulliexperimenten, d.h. $X = \sum_{i=1}^{n} X_i$

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] \overset{\uparrow}{=} n \cdot p$$

s. Bernoulli-Verteilung

$$Var[X] = n \cdot p \cdot (1-p)$$

R-Befehle:

Wahrscheinlichkeits- bzw. Dichtefkt.
d binom$(x,n,p) = P_{B_{n,p}}(x)$

Verteilungsfkt.
p binom$(x,n,p) = F_{B_{n,p}}(x)$

Quantilsfkt.
q binom$(q,n,p) = x_q$

r binom(m,n,p) : erzeugt m binomialverteilte Zufallszahlen

Beispiel: Ein Kommunikationsnetz mit n voneinander unabhängig arbeitenden Komponenten ist funktionsfähig, wenn mind. die Hälfte der Komponenten funktioniert. Die Wahrscheinlichkeit, dass eine Komponente funktioniert sei $p = 0.1$.

(a) Berechnen Sie jeweils die Wahrscheinlichkeit, dass ein Netz mit 5 bzw. 3 Komponenten funktionsfähig ist.

$\longrightarrow$ mit R möglich

(b) Für welche Werte von p ist die Wahrscheinlichkeit, dass ein System mit 5 Komponenten funktioniert größer als bei einem System mit 3 Komponenten?

$\longrightarrow$ nicht mit R möglich

Lösung: $n = 5$

(a)
$$\underbrace{P(X \geq 3)}_{\text{mind. 3 Komponenten funktionieren}} = 1 - \underbrace{\overset{P(X<3)}{P(X \leq 2)}}_{F(2)} = 1 - \text{pbinom}(2, \overset{n}{5}, \overset{p}{0.1})$$
$$= 0.00856$$

$n = 3$

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - \text{pbinom}(1, 3, 0.1) = 0.028$$

(b)
$$P_{n=5}(X \geq 3) > P_{n=3}(X \geq 2)$$
$$\sum_{x=3}^{5} \binom{5}{x} p^x (1-p)^{5-x} > \sum_{x=2}^{3} \binom{3}{x} p^x (1-p)^{3-x}$$
$$\Leftrightarrow \quad 10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3$$

$\vdots$ polyonomiale Ungleichung

$$\Rightarrow \quad p > \frac{1}{2}$$

4.2 Hypergeometrische Verteilung

Situation: n-maliges Ziehen ohne Zurücklegen aus einer endlichen Menge, deren Elemente sich in zwei Typen einteilen lassen (Bsp.: defekt / nicht defekt, männlich / weiblich)

Man beobachtet die Anzahl der Elemente vom Typ 1 in der Stichprobe mit Größe n, wobei die Gesamtmenge
M Elemente vom Typ 1 und
N Elemente vom Typ 2 umfasst.

Man sagt: Die ZV X ("Anzahl der Elemente vom Typ 1") mit den Werten $0, 1, \dots, \min(n, M)$ ist hypergeometrisch verteilt.

Verteilung: $X \sim H_{M,N,n}$ mit M: Elemente vom Typ 1
N: Elemente vom Typ 2
n: Größe der Stichprobe

Wahrscheinlichkeitsfkt.:

$$p_{H_{M,N,n}}(x) = \frac{\binom{M}{x} \cdot \binom{N}{n-x}}{\binom{M+N}{n}} \quad \text{für } x \in \{0, 1, \dots, \min(n, M)\}$$

Erwartungswert und Varianz:

$$E[X] = n \cdot \frac{M}{M+N} \approx p$$

$$Var[X] = n \cdot \frac{M}{M+N} \left(1 - \frac{M}{M+N}\right) \cdot \underbrace{\left(\frac{M+N-n}{M+N-1}\right)}_{\text{Korrekturfaktor} \approx 1}$$

Korrekturfaktor ≈ 1, falls $M+N$ sehr groß im Vergleich zu n

In diesem Fall besteht kaum ein Unterschied zw. Ziehen mit bzw. ohne Zurücklegen.

R-Funktionen: $P_{H_{M,N,n}}(x) = \text{dhyper}(x, M, N, n)$

$$P(X \leq x) = \text{phyper}(x, M, N, n)$$

Approximation durch Binomialverteilung:

Falls die Erfolgswahrscheinlichkeit der Binomialverteilung $p = \frac{M}{M+N}$, dann stimmen die Erwartungswerte für beide Verteilungen überein und die Varianzen unterscheiden sich nur durch den Korrekturfaktor, der ungefähr 1 ist, wenn $M+N$ sehr viel größer als der Stichprobenumfang n ist. Dann macht es keinen großen Unterschied, ob die Stichprobe durch Ziehen ohne oder mit Zurücklegen gewonnen wird und die hypergeometrische Verteilung lässt sich durch die (einfachere) Binomialverteilung mit Erfolgswahrscheinlichkeit $p = \frac{M}{M+N}$ approximieren.

Faustregel für die "Zulässigkeit" dieser Approximation:

$$\frac{n}{N+M} \leq 0.05 \quad \text{und} \quad M, N \text{ groß}$$

Beispiel: Für eine Charge von 100 Glühbirnen soll anhand einer Stichprobe vom Umfang $n = 22$ entschieden werden, ob die Charge angenommen wird. Befinden sich in der Stichprobe mehr als 2 defekte Glühbirnen, so wird die Lieferung abgelehnt. Die Defektrate in der Charge sei $p = 0.1$. Wie groß ist die Wahrscheinlichkeit, dass die Charge angenommen wird?

Treffer: defekte Glühbirne

Ziehen ohne Zurücklegen: $M = 10$, $N = 90$, $n = 22$

Ges.: $P(X \leq 2) = \text{phyper}(2, 10, 90, 22) \approx 61{,}7\,\%$