8. Bernoulliversuche, Binomialverteilung.

Urnenmodell:
Ziehen ohne Zurücklegen führt auf die bereits behandelte **Hypergeometrische Verteilung**.

Ziehen mit Zurücklegen führt auf die sogenannte **Binomialverteilung**, die in diesem Kapitel besprochen wird.

Zufallsexperimente, bei denen genau zwei Ergebnisse (1 für Erfolg bzw. 0 für Fehlschlag) möglich sind, heissen **Bernoulliexperimente.** Die Erfolgswahrscheinlichkeit sei $p \le 0.5$, die Wahrscheinlichkeit für Fehlschlag sei $q = 1 - p$.

Wir fragen nach der Wahrscheinlichkeit $P_n(x)$ bei einem Bernoulliversuch in n Wiederholungen genau x Erfolge zu erzielen?

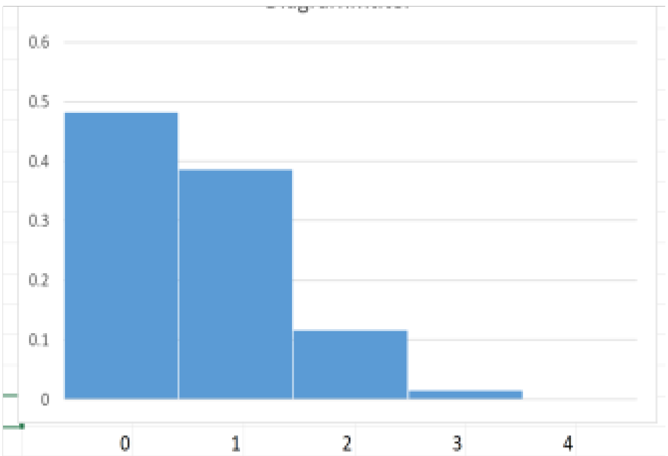
Einführendes Beispiel:
Ein Laplacewürfel wird viermal geworfen. Wie gross ist die Wahrscheinlichkeit $P_4(x)$, genau x mal eine Sechs zu würfeln?

Das Ereignis "genau zwei Sechser" {1100, 1010, 1001, 0110, 0101, 0011}
kann auf genau $\binom{4}{2}$ Arten eintreten, denn die zwei Plätze für die Augenzahl 6 (z.B. die 1. und die 4. Stelle, codiert durch 1001) können auf genau $\binom{4}{2}$ Arten ausgewählt werden.

Damit gilt: $P_4(2) = \binom{4}{2} \cdot \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2$ oder allgemeiner: $P_4(x) = \binom{4}{x} \cdot \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{4-x}$

n = 4 $\quad$ $p = \frac{1}{6}$ $\quad$ $q = \frac{5}{6}$5/6

x	$P_4(x)$
0	48.2
1	38.6
2	11.6
3	1.5
4	0.0



Die Anzahl x der Erfolge kann in diesem Beispiel die ganzzahligen Werte 0 bis 4 mit den Wahrscheinlichkeiten $P_4(x)$ annehmen. Wir sagen: Die Anzahl der Erfolge ist eine binomialverteilte Zufallsvariable. Die Wahrscheinlichkeitsverteilung ist in der Skizze grafisch dargestellt. **Zufallsvariable** werden im nächsten Kapitel besprochen.

Allgemein gilt:

Binomialverteilung

$$P_n(x) = \binom{n}{x} p^x q^{n-x} \qquad x = 0,1,...,n$$

$P_n(x)$	Wahrscheinlichkeit, bei einem Bernoulliversuch in n Wiederholungen genau x Erfolge zu erzielen.
p	Wahrscheinlichkeit für Erfolg
$q = 1 - p$	Wahrscheinlichkeit für Fehlschlag

Der Name **Binomialverteilung** weist auf den Zusammenhang mit dem Binomischen Lehrsatz hin:

$$(p+q)^2 = p^2 + 2pq + q^2 = 1$$

Die Binomialverteilung kann mit dem sogenannten **Galtonbrett** schön veranschaulicht werden. Dazu sind viele Videos im Internet zu finden.

Rekursive Berechnung der Binomialverteilung

Wegen

$$P_n(x-1) = \binom{n}{x-1} p^{x-1} q^{n-x+1} \qquad P_n(x) = \binom{n}{x} p^{x-1} \cdot p \cdot q^{n-x+1} \cdot \frac{1}{q} = \binom{n}{x} p^{x-1} \cdot q^{n-x+1} \cdot \frac{p}{q}$$

und

$$\binom{n}{x} = \frac{n(n-1)(n-2)....(n-x+2) \cdot (n-x+1)}{(x-1)! \qquad \cdot x} = \binom{n}{x-1} \cdot \frac{n-x+1}{x}$$

ergibt sich der folgende rekursive Zusammenhang:

$$P_n(x) = \frac{n-x+1}{x} \cdot \frac{p}{q} \cdot P_n(x-1)$$

Dies erlaubt mit dem Startwert $P_n(0) = q^n$ eine rasche numerische Berechnung.

$P_n(x)$ ist wachsen für x ≤ (n+1)·p und fallend für x ≥(n+1)·p. Auf beiden Seiten des Maximums fällt $P_n(x)$ schneller als exponentiell ab.

Beispiele

Geburtstage:

Am ersten Schultag werden 206 neue Schülerinnen und Schüler eingeschult- Wie gross ist die Wahrscheinlichkeit, dass kein oder 1 Geburtstagkind dabei ist oder sogar 2, 3, 4, 5?

p = 0.003 und damit q = 0.59973 und n = 206

Für die Anzahl x der Geburtstagskinder erhält man durch Rekursion die folgenden Wahrscheinlichkeiten

x	$P_n(x)$
0	0.5683
1	0.3216
2	0.0906
3	0.0169
4	0.0024
5	0.0003

Geburtenstatistik

Geissler untersuchte 1893 n = 120 137 Familien mit 4 Kindern davon x Mädchen.

x	Anzahl Familien	rel. Häufigkeit	Wahrscheinlichkeit
0	8628	**7.2**	**7.0**
1	31611	**26.3**	**26.4**
2	44793	**37.3**	**37.4**
3	28101	**23.4**	**23.6**
4	7004	**5.8**	**5.6**
5			

Anzahl Kinder: 4·120137 = 480548
Anzahl Mädchen: 0·7004 + 1·28101 + 2·44793 + 3·31611 + 4·8628 = 233516
Schätzwert für die Wahrscheinlichkeit einer Mädchengeburt p = 0.4859

Mit diesem Schätzwert ergeben sich in der Tabelle aufgeführten Wahrscheinlichkeiten:

$$P_4(x) = \binom{4}{x} \cdot p^x q^{4-x}$$

p = 0.4859 und damit q = 0.5141 und n = 4

x	0	1	2	3	4
$P_n(x)$	0.070	0.264	0.374	0.236	0.056
Aus den Daten	0.072	0.263	0.373	0.234	0.058.

Die beobachteten relativen Häufigkeiten stimmen recht gut mit den Wahrscheinlichkeiten überein. Die Güte der Übereinstimmung der Daten mit dem Binomialmodell kann mit dem sogenannten Chiquadrat-Test geprüft werden (siehe dazu das Kapitel Beurteilende Statistik).

Overbooking
Fluggesellschaften nehmen normalerweise für Linienflüge mehr Reservationen entgegen, als das Flugzeug Plätze aufweist (Überbuchung, Overbooking), denn sie rechnen durchschnittlich mit 10% Annulationen (dieser Anteil schwankt sehr stark und ist abhängig vom Flug, Zielort, Passagierklasse und Reisesaison). Wie gross ist die Wahrscheinlichkeit, dass die Gesellschaft bei einem 50-plätzigen Flugzeug Kunden zurückweisen (Kosten für Verpflegung und Unterkunft werden offenbar übernommen) oder umbuchen muss, wenn sie 52 Reservationen entgegengenommen hat?

Es geht schief, wenn von den 52 angemeldeten Personen alle oder genau 51 erscheinen:

$$P_{20}(X \geq 51) = \binom{52}{1} \cdot 0.1 \cdot 0.9^{51} + 0.9^{52} = 0.0283$$

In etwa 3 von 100 Fällen müssen Personen umgebucht werden.

KLM: 14% Annulationen (SAS Communications, 4/1992)
Swissair: Auslastungsgrad 1997: 70.5%