

Norbert Kusolitsch *

Ergänzungen zur Maß- und Wahrscheinlichkeitstheorie

7. Mai 2013

* ©2011, Institut für Statistik und Wahrscheinlichkeitstheorie der TU Wien. Jede Form der Vervielfältigung dieses Buches oder von Teilen dieses Buches ohne schriftliches Einverständnis des Autors ist strengstens untersagt.

Inhaltsverzeichnis

| | | |
|----------|--|----|
| 1 | Laplace-Räume und Kombinatorik | 1 |
| 1.1 | Laplace-Räume | 1 |
| 1.2 | Geordnete Stichproben mit Wiederholungen | 2 |
| 1.3 | Geordnete Stichproben ohne Wiederholungen | 3 |
| 1.4 | Ungeordnete Stichproben ohne Wiederholungen | 4 |
| 1.5 | ungeordnete Stichproben mit Wiederholungen | 6 |
| 2 | Ein Beispiel zum Lemma von Borel-Cantelli | 9 |
| 3 | Diskrete Wahrscheinlichkeitsverteilungen | 11 |
| 3.1 | Die hypergeometrische Verteilung | 11 |
| 3.2 | Zufallsvariable und induzierte Verteilung | 12 |
| 3.3 | Das Maximum-Likelihood-Prinzip | 13 |
| 4 | Bemerkungen zum Kapitel Unabhängigkeit | 15 |
| 4.1 | Unabhängig und paarweise unabhängig | 15 |
| 4.2 | Verallgemeinerte Multiplikationsregel | 15 |
| 4.3 | Der exakte Fisher Test | 16 |
| 4.4 | Beispiel zum Satz von der vollständigen Wahrscheinlichkeit ... | 19 |
| 4.5 | Beispiel zur Unabhängigkeit von Zufallsvariablen | 20 |
| 5 | Folgen unabhängiger Bernoulliversuche | 23 |
| 5.1 | Die Binomialverteilung | 23 |
| 5.2 | Die Multinomialverteilung B_n, p_1, \dots, p_k | 24 |
| 5.3 | Binomialverteilung versus hypergeometrische Verteilung | 25 |
| 5.4 | Die Poissonverteilung | 27 |
| 5.5 | Die geometrische Verteilung | 31 |
| 5.6 | Die negative Binomialverteilung | 32 |

Laplace-Räume und Kombinatorik

1.1 Laplace-Räume

Nachdem wir die grundlegenden Eigenschaften von Maßen kennengelernt haben, wollen wir diese auf einfache Fragestellungen aus der Wahrscheinlichkeitstheorie anwenden. Dazu benötigen wir ein paar Begriffe, die im Buch erst später behandelt werden.

Definition 1.1. Ein Tupel (Ω, \mathfrak{S}) bestehend aus einer nichtleeren Menge Ω und einer σ -Algebra \mathfrak{S} auf Ω wird Messraum genannt.

Definition 1.2. Unter einem Maßraum versteht man ein Tripel $(\Omega, \mathfrak{S}, \mu)$, bei dem (Ω, \mathfrak{S}) ein Messraum ist und μ ein Maß auf \mathfrak{S} . Ist μ ein Wahrscheinlichkeitsmaß, so spricht man von einem Wahrscheinlichkeitsraum.

Die Mengen aus \mathfrak{S} werden allgemein als messbare Mengen bezeichnet, aber bei Wahrscheinlichkeitsräumen nennt man diese Mengen Ereignisse. Die leere Menge \emptyset heißt dann unmögliches Ereignis, der ganze Raum Ω sicheres Ereignis, die einelementigen Mengen $\{\omega\} \in \mathfrak{S}, \omega \in \Omega$ nennt man in diesem Fall Elementarereignisse und die einzelnen Punkte $\omega \in \Omega$ bezeichnet man als Ausgänge oder Ergebnisse.

Ihren historischen Ausgangspunkt hat die Wahrscheinlichkeitsrechnung bei der Betrachtung von Glücksspielen genommen, bei denen nur endlich viele Ausgänge auftreten, die als gleichwahrscheinlich vorausgesetzt werden. Die dafür adäquaten Modelle bezeichnet man heute als Laplace-Räume.

Definition 1.3. Ein Laplace-Raum ist ein Wahrscheinlichkeitsraum $(\Omega, \mathfrak{S}, P)$ mit $|\Omega| < \infty, P(\{\omega\}) = P(\{\omega'\}) \quad \forall \omega, \omega' \in \Omega$ und $\mathfrak{S} = \mathfrak{P}(\Omega)$. P nennt man die diskrete Gleichverteilung auf Ω (i.Z. $U_{|\Omega|}$).

Lemma 1.4 (klassische Wahrscheinlichkeitsdefinition).

Ist $(\Omega, \mathfrak{S}, P)$ ein Laplace-Raum, so gilt für jedes $A \subseteq \Omega$

$$P(A) = \frac{|A|}{|\Omega|}.$$

Insbesondere gilt für die Elementarereignisse

$$P(\{\omega\}) = \frac{1}{|\Omega|} \quad \forall \omega \in \Omega.$$

Beweis. Mit $m := |\Omega|$ und $c := P(\{\omega\}) \quad \forall \omega$ folgt aus der Additivität von P

$$1 = P(\Omega) = \sum_{\omega \in \Omega} P(\{\omega\}) = m c \Rightarrow c = \frac{1}{m} = \frac{1}{|\Omega|}.$$

Besteht $A \subseteq \Omega$ aus den Ergebnissen $\{\omega_1, \dots, \omega_a\}$, so folgt genauso

$$P(A) = \sum_{j=1}^a P(\{\omega_j\}) = \frac{a}{m} = \frac{|A|}{|\Omega|}.$$

□

Um Wahrscheinlichkeiten in einem Laplace-Raum zu bestimmen, muss man die Anzahl der Ausgänge eines Ereignisses abzählen. Wir wollen daher kurz ein paar elementare Begriffe und Resultate der Kombinatorik vorstellen.

1.2 Geordnete Stichproben mit Wiederholungen

Als erstes betrachten wir k Ziehungen mit Zurücklegen aus $\{1, \dots, n\}$ unter Beachtung der Reihenfolge. Die Ergebnisse dieses Versuchs nennt man k -elementige geordnete Stichproben mit Wiederholungen aus $\{1, \dots, n\}$ oder kurz k -Variationen aus n .

Die Menge Ω dieser Ausgänge kann angeschrieben werden als

$$\Omega = \tilde{V}_k^n = \{(x_1, \dots, x_k) : x_i \in \{1, \dots, n\}\}.$$

Da ein k -Tupel (x_1, \dots, x_k) nur eine andere Schreibweise für eine Funktion $\varphi : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ mit $\varphi(i) = x_i$ ist, kann man \tilde{V}_k^n als Menge aller Funktionen $\varphi : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ auffassen.

Klarerweise gilt $|\tilde{V}_1^n| = n$, und, da man jedes $\hat{\varphi} : \{1, \dots, k-1\} \rightarrow \{1, \dots, n\}$ durch $\varphi(i) := \hat{\varphi}(i)$, $1 \leq i \leq k-1$ und $\varphi(k)$ aus $\{1, \dots, n\}$ beliebig auf n Arten zu einer Funktion $\varphi : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ fortsetzen kann, folgt aus der Annahme $|\tilde{V}_{k-1}^n| = n^{k-1}$ sofort $|\tilde{V}_k^n| = n^k$. Damit ist $|\tilde{V}_{k-1}^n| = n^k$ durch vollständige Induktion gezeigt.

Beispiel 1.5 (maximale Augenzahl bei 5-maligem Würfeln). Würfelt man 5-mal mit unterscheidbaren Würfeln, so kann man den Merkmalraum beschreiben

durch $\Omega = \{(x_1, \dots, x_5) : x_i \in \{1, \dots, 6\}\}$, und es gilt $|\Omega| = 6^5$. Uns interessiert die Wahrscheinlichkeit, mit der die maximale Augenzahl bei diesen Würfeln die Werte $1, \dots, 6$ annimmt, d.h. wir müssen die Wahrscheinlichkeit der Ereignisse $A_j := \{(x_1, \dots, x_5) \in \Omega : \max_{1 \leq i \leq 5} x_i = j\}$ berechnen.

Dazu betrachten wir zunächst die Ereignisse

$$B_j := \{(x_1, \dots, x_5) \in \Omega : \max x_i \leq j\} = \{(x_1, \dots, x_5) : x_i \in \{1, \dots, j\}\}.$$

Das sind die 5-Variationen aus j mit Wiederholungen. Deshalb gilt

$$|B_j| = j^5 \quad \text{und} \quad P(B_j) = \left(\frac{j}{6}\right)^5.$$

Aus $B_j \subseteq B_{j+1}$ für $i = 1, \dots, 5$, $A_j = B_j \setminus B_{j-1}$ für $2 \leq j \leq 6$ bzw. $A_1 = B_1$ folgt nun wegen der Subtraktivität der Maßfunktion

$$P(A_1) = \frac{1}{6^5}, \quad P(A_j) = P(B_j) - P(B_{j-1}) = \frac{j^5 - (j-1)^5}{6^5}, \quad 2 \leq j \leq 6.$$

Würfelt man n -mal, so gilt für das Ereignis B_5 , dass dabei kein „6-er“ auftritt, offensichtlich $P(B_5) = \left(\frac{5}{6}\right)^n$. Betrachtet man nun eine Folge von Würfeln, so ist $B := \bigcap_{n \in \mathbb{N}} B_n$, das Ereignis, dass in der gesamten Folge keine „6“ auftritt.

Man beachte, dass $B \neq \emptyset$, da $(1, 1, \dots) \in B$. Aber es gilt auch $B_n \supseteq B_{n+1} \forall n \in \mathbb{N}$, sodass aus der Stetigkeit von oben folgt

$$P(B) = \lim_n P(B_n) = \lim_n \left(\frac{5}{6}\right)^n = 0.$$

1.3 Geordnete Stichproben ohne Wiederholungen

Betrachten wir als zweiten Fall k Ziehungen aus $\{1, \dots, n\}$ ohne Zurücklegen und unter Beachtung der Reihenfolge. Das Ergebnis dieses Versuchs wird als k -elementige geordnete Stichprobe aus $\{1, \dots, n\}$ ohne Wiederholungen oder als k -Variation aus n ohne Wiederholungen bezeichnet. Man kann diesen Versuch durch das Laplace-Modell $(\Omega, \mathfrak{P}(\Omega), U_{|\Omega|})$ beschreiben mit

$$\Omega := V_k^n := \{(x_1, \dots, x_k) : x_i \in \{1, \dots, n\}, \text{ und } x_i \neq x_j \quad \forall i \neq j\}.$$

Die k -Tupel (x_1, \dots, x_k) in V_k^n entsprechen natürlich den injektiven Funktionen $\varphi : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ mit $\varphi(i) = x_i \quad \forall 1 \leq i \leq k$.

Es bleibt nur noch die Anzahl der Ausgänge zu berechnen.

Natürlich gilt $|V_1^n| = n$, und mit der Bezeichnung $(n)_k := \prod_{i=0}^{k-1} (n-i)$ folgt aus

der Annahme $|V_{k-1}^n| = (n)_{k-1}$ sofort $|V_k^n| = (n)_k$, denn jede Funktion $\hat{\varphi}$ aus V_{k-1}^n kann durch $\varphi(i) := \hat{\varphi}(i) \quad \forall 1 \leq i \leq k-1$ und mit beliebigem $\varphi(k)$ aus

$\{1, \dots, n\} \setminus \{\varphi(1), \dots, \varphi(k-1)\}$ auf $n - (k-1) = n - k + 1$ Arten zu einer Funktion φ aus V_k^n fortgesetzt werden. Somit haben wir durch Induktion nach k bewiesen, dass $|V_k^n| = \binom{n}{k}$ gilt. Für $k = n$ erhält man $|V_n^n| = \binom{n}{n} = n!$ und die Elemente von V_n^n sind die Permutationen von $\{1, \dots, n\}$.

Beispiel 1.6. Man bestimme die Wahrscheinlichkeit, dass eine zufällig gewählte Permutation π_1, \dots, π_n mindestens einen Fixpunkt enthält, sodass für ein $1 \leq i \leq n$ gilt $i = \pi_i$.

Lösung: Ist $(\pi_1, \dots, \pi_{i-1}, i, \pi_{i+1}, \dots, \pi_n)$ eine Permutation mit $i = \pi_i$, so ist $(\pi_1, \dots, \pi_{i-1}, \pi_{i+1}, \dots, \pi_n)$ eine Permutation von $(1, \dots, i-1, i+1, \dots, n)$ und umgekehrt. Daher enthält die Menge A_i der Permutationen mit $i = \pi_i$ genau $(n-1)!$ Elemente, und $A := \bigcup_{i=1}^n A_i$ ist die Menge aller Permutationen mit mindestens einem Fixpunkt. Nun folgt aus dem Additionssatz

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P\left(\bigcap_{j=1}^k A_{i_j}\right).$$

Aber die Wahrscheinlichkeiten der Durchschnitte $\bigcap_{j=1}^n A_{i_j}$ sind leicht zu bestimmen, denn $\bigcap_{j=1}^n A_{i_j}$ enthält gerade die Permutationen mit $\pi_{i_j} = i_j$, bei denen

nur die $n-k$ Zahlen aus $\{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$ permutiert werden können. Daraus folgt $\left|\bigcap_{j=1}^k A_{i_j}\right| = (n-k)!$ und $P\left(\bigcap_{j=1}^k A_{i_j}\right) = \frac{(n-k)!}{n!}$, und weiters

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{1}{\binom{n}{k}} = \sum_{k=1}^n \frac{(-1)^{k-1}}{\binom{n}{k}} \sum_{1 \leq i_1 < \dots < i_k \leq n} 1. \quad (1.1)$$

Der inneren Summe in (1.1) entspricht die Anzahl der Möglichkeiten eine k -elementige Teilmenge aus $\{1, \dots, n\}$ auszuwählen. Das ist gerade die Anzahl von Möglichkeiten k Ziehungen ohne Zurücklegen aus $\{1, \dots, n\}$ durchzuführen, wobei die Reihenfolge unerheblich ist. Damit werden wir uns im nächsten Abschnitt beschäftigen.

1.4 Ungeordnete Stichproben ohne Wiederholungen

Das Ergebnis von k Ziehungen aus $\{1, \dots, n\}$ ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge nennt man eine k -elementige, ungeordnete

Stichproben aus $\{1, \dots, n\}$ ohne Wiederholungen oder eine k -Kombination aus n ohne Wiederholungen. Für den Ergebnisraum Ω verwenden wir die Bezeichnung K_k^n , d.h. $K_k^n = \{\{x_1, \dots, x_k\} \subseteq \{1, \dots, n\}\}$.

Zu jeder Menge $\{x_1, \dots, x_k\}$ gibt es genau $k!$ geordnete Tupel $(x_{\pi_1}, \dots, x_{\pi_k})$, die durch Permutation von $1, \dots, k$ aus $\{x_1, \dots, x_k\}$ hervorgehen, und aus $\{x_1, \dots, x_k\} \neq \{y_1, \dots, y_k\}$ folgt $(x_{\pi_1}, \dots, x_{\pi_k}) \neq (y_{\tilde{\pi}_1}, \dots, y_{\tilde{\pi}_k})$ für alle zugehörigen Variationen. Also gilt $|K_k^n| k! = |V_k^n|$ bzw.

$$|K_k^n| = \frac{|V_k^n|}{k!} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}. \tag{1.2}$$

Eingesetzt in (1.1) ergibt das $P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k-1} \frac{\binom{n}{k}}{\binom{n}{k}} = \sum_{k=1}^n \frac{(-1)^{k-1}}{k!}$, und die Wahrscheinlichkeit des komplementären Ereignisses, dass eine Permutation keinen Fixpunkt besitzt, ergibt sich zu

$$P\left(\bigcap_{i=1}^n A_i^c\right) = \sum_{k=0}^n \frac{(-1)^k}{k!} \approx e^{-1}.$$

Die explizite Formel $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ ist bei großem n und k zur numerischen Berechnung ungeeignet. Dafür greift man besser auf eine Rekursion zurück, die einen anderen Weg zur Bestimmung von $|K_k^n|$ aufzeigt.

Man kann $\{x_1, \dots, x_k\}$ entweder aus $\{1, \dots, n-1\}$ auf $|K_k^{n-1}|$ Arten wählen oder man wählt $\{x_1, \dots, x_{k-1}\}$ aus $\{1, \dots, n-1\}$ auf $|K_{k-1}^{n-1}|$ Arten aus und fügt n als k -tes Element dazu. Daher gilt

$$|K_k^n| = |K_{k-1}^{n-1}| + |K_k^{n-1}|. \tag{1.3}$$

Mit $K_0^n := 1$ ergibt sich folgendes Rekursionsschema

| | | | | | |
|----------|----------|----------|----------|----------|--|
| $n = 0$ | | | | | |
| $n = 1$ | | 1 | 1 | | |
| $n = 2$ | | 1 | 2 | 1 | |
| $n = 3$ | 1 | 3 | 3 | 1 | |
| \vdots | \vdots | \vdots | \vdots | \vdots | |

das als Pascal'sches Dreieck bekannt ist.

Tatsächlich gilt $\binom{0}{0} = 1 = |K_0^0|$, $\binom{1}{0} = 1 = |K_0^1|$, $\binom{1}{1} = 1 = |K_1^1|$, und aus $\binom{n-1}{k} = |K_k^{n-1}|$ folgt

$$\begin{aligned} |K_k^n| &= |K_{k-1}^{n-1}| + |K_k^{n-1}| = \binom{n-1}{k-1} + \binom{n-1}{k} \\ &= \frac{(n-1)!}{(k-1)!(n-k-1)!} \left(\frac{1}{n-k} + \frac{1}{k} \right) = \frac{n!}{k!(n-k)!} = \binom{n}{k}. \end{aligned}$$

Also gilt $|K_k^n| = \binom{n}{k}$ für alle $n \in \mathbb{N}_0$, $0 \leq k \leq n$.

Nun gibt es zu jeder Menge $\{x_1, \dots, x_k\}$ genau eine Permutation π_1, \dots, π_k , sodass $x_{\pi_1} < \dots < x_{\pi_k}$. (In der Statistik wird für $x_{\pi_1} < \dots < x_{\pi_k}$ auch die Notation $x_{(1)} < \dots < x_{(k)}$ verwendet.)

Aber $x_{\pi_1} < \dots < x_{\pi_k}$ entspricht einer strikt monoton wachsenden Funktion $\varphi : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ mit $\varphi(i) := x_{\pi_i}$, und umgekehrt gehört zu jeder derartigen Funktion gerade eine Menge $\{\varphi(1), \dots, \varphi(k)\} \subseteq \{1, \dots, n\}$.

Man kann daher K_k^n interpretieren als die Menge der strikt monoton wachsenden Funktionen von $\{1, \dots, k\}$ in $\{1, \dots, n\}$.

Den Kombinationen $x := \{x_1, \dots, x_k\}$ aus K_k^n kann man auf bijektive Weise

Funktionen $b : \{1, \dots, n\} \rightarrow \{0, 1\}$ zuordnen durch $b^x := \begin{cases} 1, & \exists j : x_j = i \\ 0, & \text{sonst} \end{cases}$,

wobei klarerweise gilt $\sum_{i=1}^n b^x(i) = \sum_{j=1}^k b^x(x_j) = k$.

Man nennt $(b^x(1), \dots, b^x(n))$ die zu x gehörigen Besetzungszahlen.

Aus $x \neq y := \{y_1, \dots, y_k\}$ folgt $\exists j : y_j \notin \{x_1, \dots, x_k\} \Rightarrow b^x(y_j) = 0$, $b^y(y_j) = 1$ und somit $b^x \neq b^y$, d.h. die Zuordnung $x \rightarrow b^x$ ist injektiv.

Umgekehrt bildet für jede Funktion $b : \{1, \dots, n\} \rightarrow \{0, 1\}$ mit $\sum_{i=1}^n b(i) = k$ die

Menge $\{i : b(i) = 1\}$ eine k -elementige Teilmenge von $\{1, \dots, n\}$. Daher ist die Zuordnung auch surjektiv.

Beispiel 1.7 (Lotto „6 aus 45“).

$$\Omega = \{\{x_1, \dots, x_6\} \subseteq \{1, \dots, 45\}\} \Rightarrow |\Omega| = \binom{45}{6} = 8145060.$$

Sind nun $\{g_1, \dots, g_6\}$ die 6 Gewinnzahlen einer Ziehung, so ist

$$A_3 := \{\{r_1, r_2, r_3\} \cup \{s_1, s_2, s_3\} : \{r_1, r_2, r_3\} \subseteq \{g_1, \dots, g_6\}, \\ \{s_1, s_2, s_3\} \subseteq \{1, \dots, 45\} \setminus \{g_1, \dots, g_6\}\}$$

die Menge aller Kombinationen aus K_6^{45} , bei denen genau 3 Zahlen richtig und 3 Zahlen falsch sind, also die Menge aller „Dreier“.

Da es $\binom{6}{3}$ Möglichkeiten gibt $\{r_1, r_2, r_3\}$ aus $\{g_1, \dots, g_6\}$ auszuwählen und $\binom{39}{3}$ Möglichkeiten $\{s_1, s_2, s_3\}$ aus $\{g_1, \dots, g_6\}^c$ zu ziehen, gilt $|A_3| = \binom{6}{3} \binom{39}{3}$.

Daraus folgt $P(A_3) = \frac{\binom{6}{3} \binom{39}{3}}{\binom{45}{6}}$.

1.5 ungeordnete Stichproben mit Wiederholungen

Das Ergebnis von k Ziehungen aus $\{1, \dots, n\}$ mit Zurücklegen ohne Beachtung der Reihenfolge, für das wir die Bezeichnung $[x_1, \dots, x_k]$ verwenden

werden, wird k -elementige, ungeordnete Stichprobe aus $\{1, \dots, n\}$ mit Wiederholungen oder k -Kombination aus n mit Wiederholungen genannt. Für die Menge dieser Kombinationen gebrauchen wir die Notation \tilde{K}_k^n .

Ist die Reihenfolge unerheblich, so unterscheidet man nicht zwischen zwei Variationen $(x_1, \dots, x_k), (y_1, \dots, y_k)$ aus \tilde{V}_k^n , wenn sie durch Permutation ineinander übergehen, wenn also für eine Permutation π_1, \dots, π_k von $1, \dots, k$ gilt $y_{\pi_i} = x_i \quad \forall i = 1, \dots, k$. Dadurch wird eine Äquivalenzrelation auf \tilde{V}_k^n gebildet, und \tilde{K}_k^n ist die Menge der Restklassen bezüglich dieser Relation.

Ist (x_1, \dots, x_k) ein Repräsentant der Restklasse $[x_1, \dots, x_k]$, so wird dieser durch die Permutation

$$\begin{aligned} \pi_1 &:= \min \left\{ i \in \{1, \dots, k\} : x_i = \min_{1 \leq j \leq k} x_j \right\}, \\ \pi_2 &:= \min \left\{ i \in \{1, \dots, k\} \setminus \{\pi_1\} : x_i = \min_{j \notin \{\pi_1\}} x_j \right\}, \\ &\vdots \\ \pi_k &:= \min \left\{ i \in \{1, \dots, k\} \setminus \{\pi_1, \dots, \pi_{k-1}\} : x_i = \min_{j \notin \{\pi_1, \dots, \pi_{k-1}\}} x_j \right\} \end{aligned}$$

der Größe nach geordnet. Da $(x_{(1)}, x_{(2)}, \dots, x_{(k)}) := (x_{\pi_1}, x_{\pi_2}, \dots, x_{\pi_k})$ durch Permutation aus (x_1, \dots, x_k) hervorgeht, liegt es in $[x_1, \dots, x_k]$.

$(x_{(1)}, x_{(2)}, \dots, x_{(k)})$ ist, wie im folgenden Lemma gezeigt, eindeutig bestimmt.

Lemma 1.8. *Gibt es zu $x_1 \leq \dots \leq x_k$ und $y_1 \leq \dots \leq y_k$ eine Permutation π von $\{1, \dots, k\}$ mit $y_i = x_{\pi_i} \quad \forall i = 1, \dots, k$, so gilt $y_i = x_i \quad \forall i = 1, \dots, k$.*

Beweis. Gibt es ein $y_h \neq x_h$, wobei wir o.B.d.A. $y_h < x_h$ annehmen, dann gilt $y_i = x_{\pi_i} \leq y_h < x_h \leq \dots \leq x_k \quad \forall 1 \leq i \leq h \Rightarrow \pi_i \notin \{h, \dots, k\} \quad \forall 1 \leq i \leq h$. Das führt zum Widerspruch $\{\pi_1, \dots, \pi_h\} \subseteq \{1, \dots, h-1\}$. \square

Dem geordneten k -Tupel $(x_{(1)}, \dots, x_{(k)})$ entspricht aber die monoton steigenden Funktion $\varphi : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ mit $\varphi(i) = x_{(i)}$. Man kann daher \tilde{K}_k^n als Menge der monotonen Funktionen φ von $\{1, \dots, k\}$ in $\{1, \dots, n\}$ ansehen. Nun kann man jeder monotonen Funktion $\varphi : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ durch $T\varphi(i) := \varphi(i) + i - 1 \quad \forall i = 1, \dots, k$ eine strikt monoton wachsende Funktion $T\varphi : \{1, \dots, k\} \rightarrow \{1, \dots, n+k-1\}$ zuordnen.

Diese Zuordnung ist injektiv, denn aus $\varphi(i) \neq \psi(i)$ folgt $T\varphi(i) \neq T\psi(i)$. Ist andererseits $\tilde{\varphi} : \{1, \dots, k\} \rightarrow \{1, \dots, n+k-1\}$ strikt monoton wachsend, so ist φ definiert durch $\varphi(i) := \tilde{\varphi}(i) - i + 1$ eine wachsende Funktion von $\{1, \dots, k\}$ in $\{1, \dots, n\}$, da $\tilde{\varphi}(i) - i + 1 \geq \tilde{\varphi}(i-1) - i + 2 = \tilde{\varphi}(i-1) - (i-1) + 1$. Somit ist die Zuordnung auch surjektiv. Daraus folgt

$$|\tilde{K}_k^n| = |K_k^{n+k-1}| = \binom{n+k-1}{k}. \tag{1.4}$$

Zu jeder Variation mit Wiederholungen $x := (x_1, \dots, x_k) \in \tilde{V}_k^n$ kann man eine Funktion $\gamma^x : \{1, \dots, n\} \rightarrow \{0, \dots, k\}$ mit $\sum_{j=1}^n \gamma^x(j) = k$ definieren

durch $\gamma^x(j) := |\{i : x_i = j\}|$. Diese Funktion verändert sich nicht durch Vertauschung der Reihenfolge der x_1, \dots, x_k . Somit wird durch jedes x der Äquivalenzklasse $[x] := [x_1, \dots, x_k]$ dieselbe Funktion γ gebildet, deren Werte $\gamma(1), \dots, \gamma(n)$ man als Besetzungszahlen der Kombination bezeichnet.

Sind $[x] \neq [y]$ zwei Kombinationen aus \tilde{K}_k^n mit $h := \min\{i : x_{(i)} \neq y_{(i)}\}$ und $b := |\{i \leq h-1 : x_{(i)} = y_{(i)}\}|$, so gilt $\gamma^x(x_{(h)}) \geq b+1$ aber $\gamma^y(x_{(h)}) \leq b$. Daraus folgt $\gamma^x \neq \gamma^y$, d.h. die Zuordnung $[x] \rightarrow \gamma^x$ ist injektiv.

Ist umgekehrt $\gamma : \{1, \dots, n\} \rightarrow \{0, \dots, k\}$ eine Funktion mit $\sum_{j=1}^n \gamma(j) = k$ und ist γ gerade an den Stellen $i_1 < \dots < i_m$ positiv, so wird durch

$$\begin{aligned} \varphi(1) &= \dots = \varphi(\gamma(i_1)) := i_1 \\ &\vdots \\ \varphi\left(\sum_{j=1}^{m-1} \gamma(i_j) + 1\right) &= \dots = \varphi\left(\sum_{j=1}^m \gamma(i_j)\right) := i_m \end{aligned}$$

eine monoton wachsende Funktion auf $\{1, \dots, k\}$ mit Werten aus $\{1, \dots, n\}$, d.h. eine Kombination aus \tilde{K}_k^n definiert, und offensichtlich gilt $\gamma^\varphi = \gamma$. Daher ist die Zuordnung auch surjektiv.

Demnach stimmt $|\tilde{K}_k^n| = \binom{n+k-1}{k}$ überein mit der Anzahl von Möglichkeiten k als Summe von n nichtnegativen Summanden $\gamma(1), \dots, \gamma(n)$ darzustellen. Diese Anzahl lässt sich auch leicht direkt herleiten:

Man denke sich die n Summanden als n Urnen, dargestellt durch $n-1$ Trennstriche, d.h. der Raum links des ersten Strichs bildet die erste Urne, der Raum zwischen erstem und zweitem Strich die zweite Urne und schließlich der Platz rechts des $n-1$ -ten Strichs die n -te Urne. Nun verteilt man k Kugeln so auf die Urnen, dass die i -te Urne gerade $\gamma(i)$ Kugeln enthält.

$$\text{OO} || | \text{O} | | \dots | \text{O} | \text{OOO}$$

Dadurch erhält man eine, wie in der obigen Abbildung gezeigte Anordnung. Kodiert man die Kugeln durch Nullen und die Striche durch Einsen, so entspricht jeder Anordnung ein $n+k-1$ -Tupel (b_1, \dots, b_{n+k-1}) mit $b_j \in \{0, 1\}$

und $\sum_{j=1}^{n+k-1} b_j = n-1$. Davon gibt es bekanntlich $\binom{n+k-1}{n-1} = \binom{n+k-1}{k}$.

Ein Beispiel zum Lemma von Borel-Cantelli

Beispiel 2.1. Man wirft eine Münze immer wieder. A_n ist das Ereignis, dass die letzten $\lceil (1+\varepsilon) \log_2 n \rceil$ Würfe bis zum n -ten Wurf einschließlich alle „Wappen“ ergeben. Man zeige, dass gilt $P(\limsup A_n) = 0$ für alle $\varepsilon > 0$. Kann die Länge der längsten Serie von aufeinanderfolgenden „Wappen“ in den ersten n Würfeln für unendlich viele n größer als $\log_2 n$ sein?

Lösung: Es gilt

$$P(A_n) = \frac{1}{2^{\lceil (1+\varepsilon) \log_2 n \rceil}} \leq \frac{1}{2^{(1+\varepsilon) \log_2 n}} = \frac{1}{n^{1+\varepsilon}} \Rightarrow \sum_n P(A_n) < \infty.$$

Somit folgt aus dem 1-ten Lemma von Borel-Cantelli

$$P(\limsup A_n) = 0.$$

Ist L_n die Länge der Wappen-Serie bis zum einschließlich n -ten Wurf (d.h. $L_n = 0$, wenn der n -te Wurf „Adler“ ergibt), so ist die obige Aussage äquivalent zu $P\left(\limsup_n \frac{L_n}{\log_2 n} \geq 1 + \varepsilon\right) = 0 \quad \forall \varepsilon > 0$. Also gilt

$$\begin{aligned} P\left(\limsup_n \frac{L_n}{\log_2 n} > 1\right) &= P\left(\bigcup_k \left[\limsup_n \frac{L_n}{\log_2 n} \geq 1 + \frac{1}{k}\right]\right) \\ &\leq \sum_k P\left(\limsup_n \frac{L_n}{\log_2 n} \geq 1 + \frac{1}{k}\right) = 0. \end{aligned}$$

Somit gibt es zu $\omega \in \Omega$ mit Wahrscheinlichkeit 1 ein $n_0 := n_0(\omega)$, sodass für alle $n \geq n_0$ gilt $\frac{L_n}{\log_2 n} \leq 1$. Ist $n \geq 2^{n_0}$, so gilt $L_i(\omega) \leq \log_2 i \leq \log_2 n \quad \forall n_0 \leq i \leq n$. Aber für alle $i < n_0$ gilt $L_i \leq i \leq n_0 \leq \log_2 n$. Somit $\max_{1 \leq i \leq n} L_i(\omega) \leq \log_2 n \quad \forall n \geq 2^{n_0}$. Also gilt

$$P\left(\limsup_n \frac{\max_{1 \leq i \leq n} L_i(\omega)}{\log_2 n} \leq 1\right) = 1.$$

Diskrete Wahrscheinlichkeitsverteilungen

3.1 Die hypergeometrische Verteilung

Wir betrachten folgende Verallgemeinerung von Beispiel 2.1:

Beispiel 3.1. Gegeben sei eine Menge von N Elementen, die wir von $1, \dots, N$ durchnummerieren und es sei angenommen, dass die Elemente $1, \dots, A$ eine Kategorie bilden und die Elemente $A + 1, \dots, N$ eine zweite Kategorie, etwa dass $1, \dots, A$ die fehlerhaften Stücke eines Gesamtbestandes von N Stücken sind und die $N - A$ Elemente $A + 1, \dots, N$ die fehlerfreien Stücke dieses Bestandes. Dann gibt es $\binom{N}{n}$ Möglichkeiten, eine n -elementige Teilmenge $\{x_1, \dots, x_n\}$, beispielsweise zum Zweck der Qualitätskontrolle, auszuwählen. Man nennt diese Teilmenge auch *Stichprobe*.

Wenn in dieser Stichprobe gerade a fehlerhafte Stücke sind, möchte man natürlich wissen, mit welcher Wahrscheinlichkeit ein derartiges Ereignis auftritt. Klarerweise muss gelten $0 \leq a \leq n \wedge A$, aber auch $n - a$, die Anzahl der fehlerfreien Stücke in der Stichprobe, kann höchstens so groß sein wie $N - A$, die Anzahl der fehlerfreien Stücke im Gesamtbestand. Daraus folgt $n - a \leq N - A \Rightarrow a \geq n - N + A$, also $0 \vee (n - N + A) \leq a \leq n \wedge A$.

Ist $G := \{1, \dots, A\}$ und $H := \{A + 1, \dots, N\}$, so ist die Menge aller Stichproben, die genau a fehlerhafte und $n - a$ gute Stücke enthalten, gegeben durch $A_a := \{\{r_1, \dots, r_a\} \cup \{s_1, \dots, s_{n-a}\} : \{r_1, \dots, r_a\} \subseteq G, \{s_1, \dots, s_{n-a}\} \subseteq H\}$. Daher gilt

$$|A_a| = \binom{A}{a} \binom{N-A}{n-a} \quad \text{bzw.} \quad P(A_a) = \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}}.$$

Um die Wahrscheinlichkeit der Mengen A_a auszurechnen, haben wir implizit den Laplace-Raum $(\Omega = \{\{x_1, \dots, x_n\} \subseteq \{1, \dots, N\}\}, \mathfrak{P}(\Omega), U_{|\Omega|})$ benutzt. Aber einen Qualitätsprüfer werden nur die Werte $0 \vee (n - N + A) \leq a \leq n \wedge A$ und die Wahrscheinlichkeiten $P(\{a\}) := U(A_a) = \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}}$, mit denen

sie auftreten, interessieren. Zur Datenreduktion und aus Gründen der Übersichtlichkeit wird er daher den Wahrscheinlichkeitsraum $(\Omega', \mathfrak{P}(\Omega'), P')$, mit $\Omega' = \{0 \vee (n - N + A), \dots, n \wedge A\}$ und $P'(\{a\}) = \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{n}{n}}$ betrachten.

Definition 3.2. Das auf $(\Omega', \mathfrak{P}(\Omega'))$ durch die obige Gleichung definierte Wahrscheinlichkeitsmaß wird hypergeometrische Verteilung (i.Z. $H_{A, N-A, n}$) genannt.

3.2 Zufallsvariable und induzierte Verteilung

Beim oben beschriebenen Übergang von Ω zu Ω' werden sämtliche Ausgänge

$\omega \in \Omega$ durch die Funktion $X(\omega) := \sum_{a=0 \vee (n-N+A)}^{n \wedge A} a \mathbb{1}_{A_a}(\omega)$ in den Bildraum

$\Omega' := \{0 \vee (n - N + A), \dots, n \wedge A\}$ abgebildet. Der Prüfer betrachtet nun eine neue Wahrscheinlichkeitsverteilung P' auf einer σ -Algebra \mathfrak{A}' auf Ω' (in unserem Fall $\mathfrak{P}(\Omega')$), indem er jedem $A' \in \mathfrak{A}'$ die Wahrscheinlichkeit des Urbilds $X^{-1}(A') = \{\omega : X(\omega) \in A'\}$ zuordnet. Dazu muss sichergestellt sein, dass man $X^{-1}(A')$ eine Wahrscheinlichkeit zuordnen kann, dass also gilt

$$X^{-1}(A') \in \mathfrak{A} \quad \forall A' \in \mathfrak{A}' .$$

Man definiert daher:

Definition 3.3. Sind $(\Omega_i, \mathfrak{A}_i)$, $i \in \{1, 2\}$ zwei Messräume, so nennt man eine Funktion $f : \Omega_1 \rightarrow \Omega_2$ $\mathfrak{A}_1 | \mathfrak{A}_2$ -messbar, wenn gilt

$$f^{-1}(A_2) \in \mathfrak{A}_1 \quad \forall A_2 \in \mathfrak{A}_2 . \quad (3.1)$$

Statt $\mathfrak{A}_1 | \mathfrak{A}_2$ -messbar schreibt man auch $f : (\Omega_1, \mathfrak{A}_1) \rightarrow (\Omega_2, \mathfrak{A}_2)$.

Ist $(\Omega_1, \mathfrak{A}_1, P)$ ein Wahrscheinlichkeitsraum und $\Omega_2 \subseteq \mathbb{R}$, so bezeichnet man die messbaren Funktionen als Zufallsvariable und bei $\Omega_2 \subseteq \mathbb{R}^k$, $k > 1$ als Zufallsvektoren und verwendet üblicherweise Großbuchstaben X, Y, \dots .

Definition 3.4. Ist $f : \Omega_1 \rightarrow \Omega_2$ und \mathfrak{A}_2 eine σ -Algebra auf Ω_2 , so nennt man $f^{-1}(\mathfrak{A}_2) := \{f^{-1}(A_2) : A_2 \in \mathfrak{A}_2\}$ die durch f erzeugte Sigmaalgebra.

Dass $f^{-1}(\mathfrak{A}_2)$ tatsächlich eine σ -Algebra ist, ist leicht zu sehen, denn es gilt $f^{-1}(\Omega_2) = \Omega_1 \in \mathfrak{A}_1$, aus $A_2 \in \mathfrak{A}_2$ folgt $f^{-1}(A_2^c) = (f^{-1}(A_2))^c \in \mathfrak{A}_1$ und $A_n \in \mathfrak{A}_2 \quad \forall n \in \mathbb{N}$ impliziert $f^{-1}(\bigcup_n A_n) = \bigcup_n f^{-1}(A_n) \in \mathfrak{A}_1$.

Satz 3.5. Ist f eine $\mathfrak{A}_1 | \mathfrak{A}_2$ -messbare Abbildung von einem Maßraum $(\Omega_1, \mathfrak{A}_1, \mu)$ in einen Messraum $(\Omega_2, \mathfrak{A}_2)$, so wird durch

$$\mu f^{-1}(A_2) := \mu(f^{-1}(A_2)) \quad \forall A_2 \in \mathfrak{A}_2 \quad (3.2)$$

ein Maß μf^{-1} auf $(\Omega_2, \mathfrak{A}_2)$ definiert. μf^{-1} ist endlich, wenn μ endlich ist, und μf^{-1} ist eine Wahrscheinlichkeitsverteilung, wenn μ eine ist.

Beweis. μf^{-1} ist natürlich nichtnegativ, und aus der Operationstreue des Urbilds (siehe Buch Lemma 2.3) folgt sofort $\mu f^{-1}(\emptyset) = \mu(\emptyset) = 0$ und $\mu f^{-1}(\Omega_2) = \mu(\Omega_1)$. Daher ist μf^{-1} endlich, wenn μ endlich ist und aus $\mu(\Omega_1) = 1$ folgt $\mu f^{-1}(\Omega_2) = 1$.

Sind die $A'_n \in \mathfrak{A}_2$ disjunkt, so sind nach Lemma 2.3 Punkt 5. auch die $f^{-1}(A'_n)$ disjunkt. Mit Punkt 4. des Lemmas folgt daraus die σ -Additivität von μf^{-1} :

$$\begin{aligned} \mu f^{-1} \left(\bigcup_n A'_n \right) &= \mu \left(f^{-1} \left(\bigcup_n A'_n \right) \right) = \mu \left(\bigcup_n f^{-1}(A'_n) \right) \\ &= \sum_n \mu(f^{-1}(A'_n)) = \sum_n \mu f^{-1}(A'_n). \end{aligned}$$

□

Definition 3.6. Ist $(\Omega_1, \mathfrak{A}_1, \mu)$ ein Maßraum, $(\Omega_2, \mathfrak{A}_2)$ ein Messraum und $f : (\Omega_1, \mathfrak{A}_1) \rightarrow (\Omega_2, \mathfrak{A}_2)$, so nennt man das durch (3.2) auf $(\Omega_2, \mathfrak{A}_2)$ definierte Maß μf^{-1} das (durch f) induzierte Maß (die induzierte Wahrscheinlichkeitsverteilung) oder einfach das Maß (die Wahrscheinlichkeitsverteilung) von f .

Bemerkung 3.7. Wenn μ σ -endlich ist, muss μf^{-1} nicht σ -endlich sein. Ist etwa $(\Omega_1, \mathfrak{A}_1, \mu) := (\mathbb{R}, \mathfrak{B}, \lambda)$, $(\Omega_2, \mathfrak{A}_2) := (\mathbb{R}, \{\emptyset, \mathbb{R}\})$, so induziert jede Funktion $f : (\Omega_1, \mathfrak{A}_1) \rightarrow (\Omega_2, \mathfrak{A}_2)$ das Maß $\lambda f^{-1}(\emptyset) = 0$, $\lambda f^{-1}(\mathbb{R}) = \infty$, das nicht σ -endlich sein kann, da \mathfrak{A}_2 keine anderen Mengen enthält.

Ein Zufallsvektor (Zufallsvariable) wird immer nach der induzierten Verteilung benannt, so ist die Funktion X aus dem nächsten Beispiel eine hypergeometrisch verteilte Zufallsvariable $X \sim H_{A,N-A,n}$.

3.3 Das Maximum-Likelihood-Prinzip

Im folgenden Beispiel wird eine statistische Fragestellung mit Hilfe der hypergeometrischen Verteilung gelöst.

Beispiel 3.8. Um den Walbestand in einem bestimmten Meeresgebiet zu schätzen, wurde folgendermaßen verfahren: Zunächst wurden A Tiere gefangen, gekennzeichnet und sodann wieder freigelassen. Nach einiger Zeit wurden wieder n Wale gefangen. Wie kann aus der Anzahl a an markierten Tieren in dieser zweiten Stichprobe auf die Größe des Gesamtbestandes N geschlossen werden?

Intuitiv wird man annehmen, dass das Verhältnis gekennzeichneter Tiere in der zweiten Stichprobe $\frac{a}{n}$ gleich ist, dem Verhältnis aller gekennzeichneter Tiere zum Gesamtbestand N , sodass gilt

$$\frac{a}{n} \approx \frac{A}{N} \Rightarrow N \approx \frac{n}{a} A.$$

Kann diese Vorgangsweise auch objektiv begründet werden?

Nun, wie wir wissen, ist die Anzahl X markierter Tiere in der zweiten Stichprobe $H_{A,N-A,n}$ verteilt, d.h. die zugehörige Verteilung P_N ist gegeben durch

$$P_N(X = a) = \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}} \quad \text{mit} \quad \max\{0, n - (N - A)\} \leq a \leq \min\{n, A\}.$$

Allerdings ist N unbekannt.

Ein sinnvoller Weg, N zu schätzen besteht darin, jenen Wert \hat{N} zu suchen, der $P_N(X = a)$ maximiert. Dazu vergleichen wir $P_N(X = a)$ mit $P_{N+1}(X = a)$.

$$\begin{aligned} P_N(X = a) > P_{N+1}(X = a) &= \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}} > \frac{\binom{A}{a} \binom{N+1-A}{n-a}}{\binom{N+1}{n}} \\ \Leftrightarrow \frac{(N-A)! n! (N-n)!}{(n-a)! N! (N-A-n+a)!} &> \frac{(N+1-A)! n! (N-n+1)!}{(n-a)! (N-A-n+a+1)! (N+1)!} \\ \Leftrightarrow 1 > \frac{(N+1-A)(N-n+1)}{(N-A-n+a+1)(N+1)} &\Leftrightarrow \frac{N+1}{N+1-n} > \frac{N+1-A}{N-A-n+a+1} \\ \Leftrightarrow 1 + \frac{n}{N+1-n} > 1 + \frac{n-a}{N+1-A-(n-a)} \\ \Leftrightarrow nN + n - nA - n^2 + na > nN + n - n^2 - aN - a + na \\ \Leftrightarrow nA < a(N+1) &\Leftrightarrow N+1 > \frac{n}{a}A. \end{aligned}$$

Daraus folgt: $P_{\lfloor \frac{n}{a}A \rfloor}(X = a) = \max_N P_N(X = a) \Rightarrow \hat{N} = \lfloor \frac{n}{a}A \rfloor$.

Die intuitiv gewonnene Schätzung ist also jene, die die Wahrscheinlichkeit des konkreten Versuchsausgangs maximiert.

Diese Methode kann auch bei anderen Problemen verwendet werden und führt i.A. zu sehr brauchbaren Schätzern. Wir definieren allgemein:

Definition 3.9. Ist auf einem Raum $(\Omega, \mathfrak{P}(\Omega))$ mit $|\Omega| \leq \aleph_0$ eine Familie von Wahrscheinlichkeitsverteilungen $\{P_\theta : \theta \in \Theta\}$ gegeben, so nennt man jede Funktion $T : \Omega \rightarrow \Theta$ eine Schätzfunktion oder einen Schätzer. Der Schätzer heißt Maximum Likelihood Schätzer, wenn gilt

$$P_{T(\omega)}(\omega) = \max_{\theta \in \Theta} P_\theta(\omega) \quad \forall \omega \in \Omega.$$

Da die Messbarkeitsbedingung (Definition 3.3) bei diskreten Wahrscheinlichkeitsräumen wegen $\mathfrak{A}_1 = \mathfrak{P}(\Omega_1)$ immer erfüllt ist, werden wir uns mit weiteren Eigenschaften messbarer Funktionen erst später beschäftigen.

Aber bevor wir weitere diskrete Verteilungen betrachten, wollen wir zunächst den Begriff der Unabhängigkeit von Ereignissen und den der bedingten Verteilung einführen.

Bemerkungen zum Kapitel Unabhängigkeit

Wenn die Unabhängigkeit bei abstrakten Modellen oft eine natürliche Eigenschaft ist, so ist sie in der Praxis oft eine vereinfachende Modellannahme. Beispielsweise wird ein in einem Text zufällig gewählter Buchstabe nicht unabhängig vom n -ten vorigen Buchstaben sein, aber bei großem n werden wir es annehmen.

4.1 Unabhängig und paarweise unabhängig

Bemerkung 4.1. Man beachte, dass sich n Ereignisse A_1, \dots, A_n durchaus gegenseitig beeinflussen können, wenn sie paarweise unabhängig sind, wenn also gilt $P(A_i \cap A_j) = P(A_i) P(A_j) \quad \forall i \neq j$, wie das folgende Beispiel zeigt.

Beispiel 4.2. $\Omega = \{1, 2, 3, 4\}$, $P(\{i\}) = \frac{1}{4} \quad \forall i \in \{1, 2, 3, 4\}$.

$A := \{1, 2\}$, $B := \{2, 3\}$, $C := \{1, 3\}$. $P(A) = P(B) = P(C) = \frac{1}{2}$.

$P(A \cap B) = P(A) P(B) = \frac{1}{4}$, $P(A \cap C) = P(A) P(C) = \frac{1}{4}$,

$P(B \cap C) = P(B) P(C) = \frac{1}{4}$.

Aber $A \cap B \cap C = \emptyset \Rightarrow 0 = P(A \cap B \cap C) \neq P(A) P(B) P(C) = \frac{1}{8}$.

Wenn man weiß, dass A und B eingetreten sind, dann ist der Ausgang „2“.

Somit ist C unmöglich.

4.2 Verallgemeinerte Multiplikationsregel

Satz 4.3. Für beliebige Ereignisse A_1, \dots, A_n auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ gilt

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \prod_{i=2}^n P\left(A_i \mid \bigcap_{j=1}^{i-1} A_j\right).$$

Beweis. Vollständige Induktion: $n = 2$:

$$P(A_1 \cap A_2) = P(A_1) \frac{P(A_1 \cap A_2)}{P(A_1)} = P(A_1)P(A_2|A_1).$$

$n - 1 \rightarrow n$:

$$\begin{aligned} P\left(\bigcap_{i=1}^n A_i\right) &= P\left(\bigcap_{i=1}^{n-1} A_i\right) P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right) \\ &= P(A_1) \prod_{j=1}^{n-1} P\left(A_j \mid \bigcap_{i=1}^{j-1} A_i\right) P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right) \\ &= P(A_1) \prod_{j=2}^n P\left(A_j \mid \bigcap_{i=1}^{j-1} A_i\right). \end{aligned}$$

□

4.3 Der exakte Fisher Test

In der Praxis weiß man oft nicht, ob Ereignisse unabhängig sind. Das folgende Verfahren hilft uns bei der Überprüfung dieser Annahme.

Man möchte prüfen, ob die Ereignisse B , C unabhängig sind. Führt man einen Versuch N -mal durch und notiert, ob B oder B^c , C oder C^c eingetreten ist, so erhält man eine sogenannte 2×2 -Kontingenztafel (Vierfeldertafel) der folgenden Gestalt:

| | B | B^c | Zeilensummen |
|--------------|----------|----------|--------------|
| C | x_{11} | x_{12} | A |
| C^c | x_{21} | x_{22} | $N - A$ |
| Spaltensumme | n | $N - n$ | N |

Dabei gibt x_{11} die Anzahl X der Versuche an, bei denen sowohl B als auch C eingetreten ist. x_{12} ist die Anzahl der Versuche, bei denen C aber nicht B eingetreten ist usw. Ist A die Anzahl der Versuche, bei denen C eingetreten ist und n die Anzahl der Versuche, bei denen das Ereignis B auftritt, so gilt weiters $A = x_{11} + x_{12}$, $N - A = x_{21} + x_{22}$, $n := x_{11} + x_{21}$ und $N - n := x_{12} + x_{22}$, d.h. $A, N - A$ sind die Zeilensummen der Tafel und $n, N - n$ ihre Spaltensummen. Sind diese Summen festgelegt, so sind durch x_{11} auch alle anderen x_{ij} bestimmt: $x_{12} = A - x_{11}$, $x_{21} = n - x_{11}$, $x_{22} = N - A - n + x_{11}$.

Der Wert der Zufallsvariablen X beschreibt also die Vierfeldertafel vollständig bei fixem N , A und n . Wenn die Annahme stimmt, dass B und C unabhängig sind (man bezeichnet diese Annahme auch als Nullhypothese), dann wird

durch den Eintritt von B eine zufällige Stichprobe von n Versuchen festgelegt und X ist die Anzahl der Versuche in dieser Stichprobe, bei denen C eintritt. Daher ist X bei Zutreffen der Nullhypothese und unter der Annahme, dass N , A und n fixiert sind $H_{A, N-A, n}$ verteilt.

Der Anteil der Versuche in der Stichprobe, bei denen C eintritt, sollte dem Anteil $\frac{A}{N}$ aller Versuche entsprechen, bei denen C aufgetreten ist. Somit

$$\frac{X}{n} \approx \frac{A}{N} \quad \text{bzw.} \quad X \approx n \frac{A}{N} \quad \Rightarrow \quad \frac{X}{N} \approx \frac{n}{N} \frac{A}{N}.$$

Bei fixierten Zeilen- und Spaltensummen gibt $\frac{A}{N}$ die Wahrscheinlichkeit von C an und $\frac{n}{N}$ ist die Wahrscheinlichkeit von B , d.h. $P(B) = \frac{n}{N}$, $P(C) = \frac{A}{N}$. $\frac{X}{N}$ entspricht der Wahrscheinlichkeit von $B \cap C$, und daher sollte bei Unabhängigkeit von B und C gelten

$$\frac{X}{N} = P(B \cap C) \approx P(B)P(C) = \frac{n}{N} \frac{A}{N}.$$

Unterschreitet X den Wert $n \frac{A}{N}$ sehr stark, so deutet dies darauf hin, dass C den Eintritt von B *behindert*, zu großes X wiederum lässt darauf schließen, dass B durch C *begünstigt* wird. Daher berechnet man die Wahrscheinlichkeit, dass $|X - n \frac{A}{N}|$ mindesten den beobachteten Wert $d := |x_{11} - n \frac{A}{N}|$ annimmt.

$$\begin{aligned} P\left(\left|X - n \frac{A}{N}\right| \geq d\right) &= P\left(X \leq n \frac{A}{N} - d\right) + P\left(X \geq n \frac{A}{N} + d\right) \\ &= \sum_{a=0 \vee (n-N+A)}^{\lfloor \frac{nA}{N} - d \rfloor} \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}} + \sum_{a=\lceil \frac{nA}{N} + d \rceil}^{n \wedge A} \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}}. \end{aligned}$$

Diese Wahrscheinlichkeit wird p -Wert genannt. Ist sie kleiner oder gleich einer vorgegebenen Schranke α , so wird die Hypothese abgelehnt. Demnach entscheidet man mit Wahrscheinlichkeit α fehlerhaft, falls die Annahme dennoch richtig ist. Aus diesem Grund nennt man α die Fehlerwahrscheinlichkeit 1. Art (falls nicht die Plausibilität einer Annahme geprüft wird, sondern wenn zwischen zwei konkreten Annahmen zu entscheiden ist, gibt es auch eine Fehlerwahrscheinlichkeit 2. Art, auf die wir hier nicht näher eingehen).

Lautet die Nullhypothese „ C behindert B “ oder „ C begünstigt B “, so spricht man von einer einseitigen Fragestellung.

Im 1-ten Fall wird die Annahme beschrieben durch $P(B \cap C) \leq P(B)P(C)$, und dem widersprechen nur große X -Werte. Klarerweise wird die Wahrscheinlichkeit $P(X \geq x_{11})$ umso geringer, je kleiner die Wahrscheinlichkeit $q := P(B \cap C)$ ist, oder anders ausgedrückt die Wahrscheinlichkeit $P(X \geq x_{11})$ wird unter dieser Nullhypothese gerade für $P(B \cap C) = P(B)P(C)$ maximal. Der p -Wert ergibt sich daher zu

$$P(X \geq x_{11}) = \sum_{a=x_{11}}^{n \wedge A} \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}}.$$

Die Hypothese wird zu einem vorgegebenen Fehlerniveau α abgelehnt, wenn diese p -Wert höchstens α ist.

Im zweiten Fall haben wir die Nullhypothese $H_0 : P(B \cap C) \geq P(B)P(C)$ und kleine X -Werte widersprechen dieser Annahme. Unter dieser Nullhypothese wird die Wahrscheinlichkeit $P(X \leq x_{11})$ maximal für $P(B \cap C) = P(B)P(C)$ und der p -Wert ergibt sich daher zu

$$P(X \leq x_{11}) = \sum_{a=0 \vee (n-N+A)}^{x_{11}} \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}}.$$

Bemerkung 4.4. Klarerweise kann man die Rollen von B und C bzw. von Zeilen und Spalten vertauschen.

Ein paar Beispiele mögen das bisher Gesagte veranschaulichen.

Beispiel 4.5. („Medikamenten Test“). Ein bisher noch nicht erprobtes Medikament wird an 10 Patienten getestet, von denen 8 geheilt werden. Von 5 Patienten, die das Medikament nicht bekommen, wird nur einer geheilt. Ist die Heilungschance unabhängig von der Medikamentenbehandlung?

Da das Medikament noch nicht erprobt wurde, ist es durchaus denkbar, dass es die Heilungschancen nicht nur nicht verbessert, sondern sogar noch verschlechtert. Es ist daher sinnvoll von einer 2-seitigen Fragestellung auszugehen. Mit $B :=$ „geheilt“ und $C :=$ „mit Medikamenten behandelt“ erhält man die Kontingenztafel

| | geheilt | nicht geheilt | |
|-----------------|---------|---------------|-------------|
| Medikament | 8 | 2 | $10 = A$ |
| kein Medikament | 1 | 4 | $5 = N - A$ |
| | $9 = n$ | $6 = N - n$ | $15 = N$ |

Hätte das Medikament keinen Einfluss auf die Heilung, so sollte X den Wert $n \frac{A}{N} = 9 \frac{10}{15} = 6$ annehmen. Daraus folgt $d = 2$ und der p -Wert ergibt sich für $X \sim H_{10,5,9}$ zu $P(X \leq 4) + P(X \geq 8) \approx 0.08891$. Die Nullhypothese ist eher unglaubwürdig.

Angenommen früher durchgeführte Tests an anderen Kliniken lassen vermuten, dass das Medikament den Krankheitsverlauf positiv beeinflusst, dann stellt man zunächst die gegenteilige Behauptung $P(B \cap C) \leq P(B)P(C)$ als Nullhypothese auf und hofft sie durch einen kleinen p -Wert in den Versuchen widerlegen zu können. Wir erhalten in diesem Fall für den p -Wert $p = P(X \geq 8) = \sum_{a=8}^9 \frac{\binom{10}{a} \binom{5}{9-a}}{\binom{15}{9}} \approx 0.04695$ und man wird daraus schließen, dass das Medikament die Heilungschancen verbessert.

Beispiel 4.6. („Fahrtüchtigkeit“). Es soll der Einfluss von Alkohol auf die Fahrtüchtigkeit getestet werden. Dazu werden 30 Testpersonen in zwei gleich

große Gruppen eingeteilt. Die erste Gruppe setzt sich nach dem Konsum einer bestimmten Menge Alkohol an einen Fahrsimulator. Die zweite Gruppe wird nüchtern getestet. Von der ersten Gruppe verursachen 12 einen Unfall, von der nüchternen Gruppe nur 4. Damit erhält man folgende 2×2 -Tafel:

| | | | |
|-----------------|--------|-------------|------------|
| | Unfall | kein Unfall | |
| Medikament | 12 | 3 | 15 = A |
| kein Medikament | 4 | 11 | 15 = N - A |
| | 16 = n | 14 = N - n | 30 = N |

Auf Grund aller bisherigen Erfahrungen wird man erwarten, dass der Alkoholkonsum die Unfallhäufigkeit steigert und man formuliert daher die gegenteilige Nullhypothese „Alkoholkonsum vermindert die Unfallhäufigkeit“. Der p -Wert für diese einseitige Fragestellung ergibt sich daher mit $X \sim H_{15,15,16}$ zu

$$P(X \geq 12) = \sum_{a=12}^{15} \frac{\binom{15}{a} \binom{15}{16-a}}{\binom{30}{16}} \approx 0.00461,$$

d.h. selbst wenn kein Zusammenhang zwischen Alkoholkonsum und Unfallhäufigkeit bestünde, wäre mit diesem Ergebnis nur mit einer Wahrscheinlichkeit von 0.00461 zu rechnen. Umso geringer wird diese Wahrscheinlichkeit sein, wenn Alkoholkonsum die Unfallgefahr steigert und man wird daher die Nullhypothese verwerfen.

4.4 Beispiel zum Satz von der vollständigen Wahrscheinlichkeit

Beispiel 4.7. Man möchte herausfinden, wie groß der Anteil p der Leute in einer bestimmten Personengruppe ist, die schon einmal Drogen genommen haben. Um ehrliche Antworten zu erhalten, vereinbart man mit den Testpersonen folgende Vorgangsweise:

Der Befragte würfelt zunächst einmal unbeaufsichtigt und schreibt dann eine falsche Antwort auf den Fragebogen, wenn der Würfel auf 6 fällt, ansonsten beantwortet er die Frage wahrheitsgemäß.

Auf diese Art kann jeder, der mit „ja“ geantwortet hat, dies mit einer „Sechs“ beim Würfeln erklären. Wenn sich alle Testpersonen an die Spielregeln halten und ein Anteil $\hat{p} = 0.2$ der Fragebogen enthält ein „ja“, wie groß ist dann p ?

Lösung: Ist W_6 das Ereignis, dass der Würfel auf 6 fällt, J das Ereignis, dass der Befragte mit „ja“ antwortet, D das Ereignis, dass die Testperson schon einmal Drogen genommen hat, so gilt

$$\hat{p} = P(J) = P(W_6)P(J|W_6) + P(W_6^c)P(J|W_6^c). \quad (4.1)$$

Klarerweise gilt $P(W_6) = \frac{1}{6}$, $P(W_6^c) = \frac{5}{6}$. Falls sich alle Personen an die vereinbarte Vorgehensweise halten, muss unter den Personen, die eine „6“ gewürfelt haben jeder, der noch nie Drogen genommen hat, mit „ja“ antworten. Daher muss gelten $P(J|W_6) = 1 - p$, und umgekehrt wird von den Personen, die keine „6“ gewürfelt haben, nur der Anteil p der tatsächlich schon Drogenerfahrungen hat, mit „ja“ antworten. Daher gilt $P(J|W_6^c) = p$. Setzt man dies in die Gleichung (4.1) ein, so erhält man

$$\hat{p} = \frac{1}{6}(1-p) + \frac{5}{6}p \Rightarrow p = \frac{6\hat{p} - 1}{4} = \frac{1}{20} = 0.05.$$

Fünf Prozent der Befragten haben also schon einmal Drogen genommen. Man beachte, dass \hat{p} nicht kleiner als $\frac{1}{6}$ sein kann, denn selbst bei $p = 0$ müssen zumindest die Personen, die eine „6“ gewürfelt haben, mit „ja“ antworten. Wenn sich nun eine Testperson fragt, wie glaubwürdig es ist, zu sagen, man habe eine „6“ gewürfelt, wenn man mit „ja“ geantwortet hat, oder anders gesagt: wie groß die bedingte Wahrscheinlichkeit $P(W_6|J)$ ist, so lässt sich auch das schnell beantworten, denn wegen des Bayes'schen Theorems gilt

$$P(W_6|J) = \frac{P(J \cap W_6)}{P(J)} = \frac{P(J|W_6)P(W_6)}{P(J)} = \frac{(1-p)\frac{1}{6}}{\hat{p}} = \frac{19}{24} \approx 0.792.$$

Ein „ja“ lässt also noch lange nicht den Schluss zu, dass die Testperson schon einmal mit Drogen in Berührung gekommen ist.

4.5 Beispiel zur Unabhängigkeit von Zufallsvariablen

Beispiel 4.8. Wir würfeln 2-mal, und beschreiben diesen Versuch durch das Laplace-Modell $(\Omega, \mathfrak{P}(\Omega), U_{|\Omega|})$ mit $\Omega = \{(x_1, x_2) : x_i \in \{1, \dots, 6\}\}$. Bezeichnet $X_i((x_1, x_2)) = x_i$ das Ergebnis des i -ten Wurfs, so gilt

$$\begin{aligned} P([X_1 = x_1] \cap [X_2 = x_2]) &= P(X_1 = x_1)P(X_2 = x_2) \\ \Rightarrow P([X_2 = x_2] | [X_1 = x_1]) &= P(X_2 = x_2) = \frac{1}{6} \quad \forall x_1, x_2. \end{aligned}$$

Die Kenntnis von X_1 sagt somit nichts über den Wert von X_2 aus.

Für $S := X_1 + X_2$ gilt wohl $P(S = 7) = P(\{(1, 6), (2, 5), \dots, (6, 1)\}) = \frac{6}{36} = \frac{1}{6}$ und $P(S = 7, X_2 = x) = P([X_2 = x] \cap [X_1 = 7 - x]) = P(S = 7)P(X_2 = x)$, woraus folgt $P(X_2 = x | S = 7) = P(X_2 = x)$, $1 \leq x \leq 6$. D.h nimmt S den Wert 7 an, so kann man nichts über X_2 aussagen.

Aber es gilt $P(S = 2) = P((1, 1)) = \frac{1}{36}$, und $P(X_2 = 1 | S = 2)$ ergibt sich zu $P(X_2 = 1 | S = 2) = \frac{P([X_1=1] \cap [S=2])}{P(S=2)} = \frac{P(X_1=1, X_2=1)}{P(X_1=1, X_2=1)} = 1$. Demnach wird X_2 durch $S = 2$ vollständig bestimmt.

Damit kein durch die Zufallsvariable X_1 bestimmtes Ereignis (das ist ein Ereignis der Form $[X_1 \in A]$, $A \in \mathfrak{B}$) Informationen über eine 2-te Zufallsvariable X_2 liefert, muss gelten

$$P([X_1 \in A] \cap [X_2 \in B]) = P([X_1 \in A]) P([X_2 \in B]) \quad \forall A, B \in \mathfrak{B}. \quad (4.2)$$

Da die Urbilder $C = [X_1 \in A]$, $A \in \mathfrak{B}$, $D = [X_2 \in B]$, $B \in \mathfrak{B}$ gerade die von den beiden Zufallsvariablen erzeugten σ -Algebren $X_1^{-1}(\mathfrak{B})$ bzw. $X_2^{-1}(\mathfrak{B})$ bilden, ist (4.2) äquivalent zur Unabhängigkeit dieser beiden σ -Algebren.

Man bezeichnet daher zwei Zufallsvariablen X_1, X_2 als unabhängig voneinander, wenn ihre zugehörigen σ -Algebren unabhängig sind.

Folgen unabhängiger Bernoulliversuche

Dieser Abschnitt beschäftigt sich mit einigen Verteilungen, die auftreten, wenn man unabhängige Wiederholungen von Alternativversuchen betrachtet.

5.1 Die Binomialverteilung

Führt man einen Alternativversuch, der mit Wahrscheinlichkeit p eine „1“ liefert, unabhängig voneinander n -mal durch, so kann man die Menge der möglichen Versuchsausgänge darstellen als

$$\Omega = \{\mathbf{x} := (x_1, \dots, x_n) : x_i \in \{0, 1\}, i = 1, \dots, n\}.$$

Allerdings sind die einzelnen Ausgänge nun nicht gleichwahrscheinlich, sondern es gilt mit $x := \sum_{i=1}^n x_i$

$$P((x_1, \dots, x_n)) = \prod_{i=1}^n [p^{x_i} (1-p)^{1-x_i}] = p^x (1-p)^{n-x}.$$

x ist natürlich die Anzahl der „Einsen“ und $n-x$ die Anzahl der „Nullen“ in den n Versuchen, und $X(\mathbf{x}) := \sum_{i=1}^n x_i$ bildet den Raum Ω ab in $\Omega' = \{0, \dots, n\}$.

Um die induzierte Verteilung PX^{-1} zu bestimmen, benötigt man die Wahrscheinlichkeiten der Urbilder $[X = x]$.

Da es $\binom{n}{k}$ Möglichkeiten gibt, x Indices aus $1, \dots, n$ für die Plätze der „Einsen“ auszuwählen, gilt $|[X = x]| = \binom{n}{x}$. Aber für jeden einzelnen Ausgang $(x_1, \dots, x_n) \in [X = x]$ gilt $P((x_1, \dots, x_n)) = p^x (1-p)^{n-x}$. Somit erhält man

$$PX^{-1}(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ mit } x = 0, \dots, n.$$

Definition 5.1 (Binomialverteilung). Eine Zufallsvariable X mit Wertebereich $\{0, \dots, n\}$ und $PX^{-1}(x) = \binom{n}{x} p^x (1-p)^{n-x}$ heißt binomialverteilt mit den Parametern n und p (i.Z. $X \sim B_{n,p}$).

Bezeichnet man mit X_i das Ergebnis des i -ten Versuchs, d.h. $X_i(\mathbf{x}) = x_i$, $i = 1, \dots, n$, so ist $X_i \sim B_p$ für $i = 1, \dots, n$ und es gilt

$$X = \sum_{i=1}^n X_i, \quad X_i \sim B_p, \quad \forall i = 1, \dots, n, \quad (5.1)$$

d.h. eine binomialverteilte Zufallsvariable X kann dargestellt werden als Summe von unabhängigen, alternativverteilten Zufallsvariablen X_i .

Man kann diese Situation durch ein Urnenmodell beschreiben:

Eine Urne enthält A weiße und $N - A$ schwarze Kugeln mit $p := \frac{A}{N}$. Man zieht n -mal aus der Urne, wobei die gezogene Kugel nach jeder Ziehung sofort wieder in die Urne zurückgegeben wird, und zählt die Anzahl X der weißen Kugeln im Verlauf der n Ziehungen. Man nennt daher die Binomialverteilung auch die *Verteilung von Ziehungen mit Zurücklegen*.

Bemerkung 5.2. Die Bernoulliverteilung kann natürlich als Sonderfall der Binomialverteilung mit dem Parameter $n = 1$ aufgefasst werden.

5.2 Die Multinomialverteilung B_n, p_1, \dots, p_k

Den oben beschriebenen Versuch kann man auf den Fall verallgemeinern, bei dem die Urne nicht zwei Arten von Kugeln enthält, sondern Kugeln in k verschiedenen Farben, wobei p_i der Anteil der Kugeln der Farbe i ist, sodass gilt

$0 \leq p_i \leq 1 \forall i = 1, \dots, k$ und $\sum_{i=1}^k p_i = 1$. Man zieht n -mal mit Zurücklegen.

Klarerweise kann man diesen Versuch beschreiben durch den Merkmalraum $\Omega := \{\mathbf{x} := (x_1, \dots, x_n) : x_i \in \{1, \dots, k\}\}$.

Die Wahrscheinlichkeit $p(x_i)$, dass bei der i -ten Ziehung Farbe x_i gezogen wird, ergibt sich mit $\delta_{i,j} := \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ ($\delta_{i,j}$ ist das Kronecker-Symbol) zu

$$p(x_i) = \prod_{j=1}^k p_j^{\delta_{j,x_i}}.$$

Da die Ziehungen mit Zurücklegen, also unabhängig erfolgen, wird man einem konkreten Ausgang (x_1, \dots, x_n) folgende Wahrscheinlichkeit zuweisen

$$P((x_1, \dots, x_n)) = \prod_{i=1}^n p(x_i) = \prod_{i=1}^n \prod_{j=1}^k p_j^{\delta_{j,x_i}} = \prod_{j=1}^k p_j^{\sum_{i=1}^n \delta_{j,x_i}}. \quad (5.2)$$

Bezeichnet man mit n_j die Anzahl der Ziehungen, bei denen eine Kugel der Farbe j gezogen wird, so gilt $n_j = \sum_{i=1}^n \delta_{j,x_i}$, und (5.2) wird zu

$$P(\mathbf{x}) = \prod_{j=1}^k p_j^{n_j} \text{ mit } 0 \leq n_j \leq n \text{ und } \sum_{j=1}^k n_j = n. \quad (5.3)$$

Durch $\mathbf{N}(\mathbf{x}) := (N_1, \dots, N_k)(\mathbf{x}) := (n_1, \dots, n_k)$ wird Ω abgebildet in den Bildraum $\Omega' = \left\{ \mathbf{n} := (n_1, \dots, n_k) : 0 \leq n_j \leq n, \sum_{j=1}^k n_j = n \right\}$.

Um darauf die induzierte Verteilung zu bestimmen benötigt man die Wahrscheinlichkeiten der Urbilder $[\mathbf{N} = \mathbf{n}]$. Jedes $\mathbf{x} \in [\mathbf{N} = \mathbf{n}]$ besitzt die durch (5.3) gegebene Wahrscheinlichkeit. Es genügt also $|\mathbf{N} = \mathbf{n}|$ zu bestimmen.

Nun gibt es $\binom{n}{n_1}$ Möglichkeiten aus den Indices $1, \dots, n$ die n_1 Ziehungen für Farbe „1“ auszuwählen. Dann gibt es $\binom{n-n_1}{n_2}$ Möglichkeiten aus den verbleibenden Indices n_2 Ziehungen für Farbe „2“ zu wählen, usw.. Schließlich erhält

man $|\mathbf{N} = \mathbf{n}| = \binom{n}{n_1} \prod_{j=2}^k \binom{n - \sum_{i=1}^{j-1} n_i}{n_j} = \frac{n!}{n_1! \dots n_k!} = \binom{n}{n_1, \dots, n_k}$. Daraus folgt

$$P\mathbf{N}^{-1}(\mathbf{n}) = \binom{n}{n_1, \dots, n_k} \prod_{j=1}^k p_j^{n_j} \text{ für } \mathbf{n} \in \Omega'.$$

Definition 5.3. Diese Verteilung wird *Multinomialverteilung* mit den Parametern n und p_1, \dots, p_k genannt.

$|\mathbf{N} = \mathbf{n}|$ kann man auch folgendermaßen herleiten:

Es gibt $n!$ Permutationen der Zahlen $1, \dots, n$. Ersetzt man in jeder Permutation die Zahlen $1, \dots, n_1$ durch ein gemeinsames Symbol s_1 , so verbleiben $\frac{n!}{n_1!}$ unterscheidbare n -Tupel, in denen die Zahlen $n_1 + 1, \dots, n$ genau einmal und das Symbol s_1 n_1 -mal vorkommt. Ersetzt man nun die Zahlen $n_1 + 1, \dots, n_1 + n_2$ durch ein Symbol s_2 , verbleiben $\frac{n!}{n_1! n_2!}$ unterscheidbare n -Tupel. Schließlich erhält man $\frac{n!}{\prod_{j=1}^k n_j!}$

Tupel mit n_1 Symbolen s_1 , n_2 Symbolen s_2, \dots, n_k Symbolen s_k .

5.3 Binomialverteilung versus hypergeometrische Verteilung

In Abschnitt 3.1 wurde die hypergeometrische Verteilung besprochen. Bei dem ihr zu Grunde gelegten Versuchsschema wird zunächst eine n -elementige Teilmenge aus N Elementen ausgewählt. Da eine derartige Auswahl durch

Ziehungen ohne Zurücklegen erfolgt, wird die hypergeometrische Verteilung auch *Verteilung der Ziehungen ohne Zurücklegen* genannt.

Definiert man nun zur i -ten Ziehung eine Zufallsvariable X_i , die den Wert 1 annimmt, wenn bei dieser Ziehung ein Element aus $\{1, \dots, A\}$ gezogen wird, und die ansonsten 0 ist, und bezeichnet man mit X die Anzahl der Elemente aus $\{1, \dots, A\}$ im Verlauf der n Ziehungen, so gilt $X \sim H_{A, N-A, n}$ und

$$X = \sum_{i=1}^n X_i. \quad (5.4)$$

Die obige Gleichung zeigt, dass man eine hypergeometrisch verteilte Zufallsvariable, ähnlich wie bei der Binomialverteilung (siehe (5.1)), als Summe von n alternativverteilten Zufallsvariablen X_i darstellen kann.

Auch bei Ziehung ohne Zurücklegen gilt klarerweise $X_1 \sim H_{\frac{A}{N}}$. Nicht ganz so offensichtlich ist, dass auch gilt $X_i \sim H_{\frac{A}{N}}$ für $i = 2, \dots, n$. Dies kann man folgendermaßen herleiten:

Man gibt alle N Elemente in eine Urne und zieht solange ohne Zurücklegen, bis die Urne leer ist. Jedesmal, wenn man ein Element aus $\{1, \dots, A\}$ zieht, schreibt man eine „Eins“ und ansonsten eine „Null“. Dadurch erhält man ein N -Tupel (x_1, \dots, x_N) , das genau A „Einsen“ und $N - A$ „Nullen“ enthält. Klarerweise gibt es $\binom{N}{A}$ derartige Tupel, und, wenn man „blind“ zieht, so muss jeder Ziehungsverlauf, und damit jedes N -Tupel die gleiche Wahrscheinlichkeit besitzen.

Es gibt aber genau $\binom{N-1}{A-1}$ Tupel, für die gilt $x_i = 1$ für ein gegebenes i und bei denen auf die restlichen $N - 1$ Plätze $A - 1$ Einsen aufgeteilt werden. Da aber gerade für die zuletzt beschriebenen N -Tupel gilt:

$$X_i = 1, \text{ folgt daraus } P(X_i = 1) = \frac{\binom{N-1}{A-1}}{\binom{N}{A}} = \frac{A}{N}.$$

Setzt man $p := \frac{A}{N}$, so gilt $X = \sum_{i=1}^n X_i$ mit $X_i \sim B_p \quad \forall i$, egal, ob X binomial- oder hypergeometrisch verteilt ist. Man beachte aber, dass die X_i bei binomialverteiltem X unabhängig sind, während bei hypergeometrisch verteiltem X Abhängigkeiten zwischen den einzelnen Ziehungen bestehen. Daher unterscheidet sich auch die gemeinsame Verteilung von (X_1, \dots, X_n) , je nachdem, ob man mit oder ohne Zurücklegen zieht.

Sind nun N und A sehr groß im Vergleich zu n , so wird man annehmen, dass es kaum eine Rolle spielt, ob man mit oder ohne Zurücklegen zieht.

Tatsächlich gilt

Satz 5.4. Aus $0 \leq A \leq N$ und $\lim_{N \rightarrow \infty} \frac{A}{N} = p$ folgt

$$\lim_{N \rightarrow \infty} \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}} = \binom{n}{a} p^a (1-p)^{n-a} \quad \forall a \in \{0, \dots, n\}. \quad (5.5)$$

Beweis.

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{\binom{A}{a} \binom{N-A}{n-a}}{\binom{N}{n}} &= \lim_{N \rightarrow \infty} \frac{n!}{a!(n-a)!} \cdot \frac{(A)_a (N-A)_{n-a}}{(N)_n} \\ &= \binom{n}{a} \lim_{N \rightarrow \infty} \prod_{i=0}^{a-1} \frac{A-i}{N-i} \prod_{j=0}^{n-a-1} \frac{N-A-j}{N-a-j} \end{aligned} \quad (5.6)$$

Aber aus $\lim_{N \rightarrow \infty} \frac{A}{N} = p$ folgt $\lim_{N \rightarrow \infty} \frac{A-i}{N-i} = \lim_{N \rightarrow \infty} \frac{\frac{A}{N} - \frac{i}{N}}{1 - \frac{i}{N}} = p \quad \forall 0 \leq i \leq a-1$
 und $\lim_{N \rightarrow \infty} \frac{N-A-j}{N-a-j} = \lim_{N \rightarrow \infty} \frac{1 - \frac{A}{N} - \frac{j}{N}}{1 - \frac{a+j}{N}} = 1 - p \quad \forall 0 \leq j \leq n-a-1$. Eingesetzt in (5.6) ergibt das (5.5). \square

5.4 Die Poissonverteilung

Oft ist n auch für eine rekursive Berechnung der Wahrscheinlichkeiten der Binomialverteilung zu groß. Wir beginnen mit einem Beispiel.

Beispiel 5.5. Ein Kuchen, der in $m = 8$ Stücke geteilt wird, enthält $n = 100$ Rosinen. Ein Kunde möchte wissen, wie groß die Wahrscheinlichkeit ist, dass das Kuchenstück, das er kauft, mindestens 7 Rosinen enthält, wenn die Rosinen gleichmäßig und rein zufällig in den Kuchenteig eingemischt werden.

Lösung: Unter obiger Annahme sind die Zufallsvariablen X_i , definiert durch

$$X_i := \begin{cases} 1, & i\text{-te Rosine im Kuchenstück des Kunden} \\ 0, & \text{sonst} \end{cases} \quad \text{unabhängig und}$$

$B_{\frac{1}{m}}$ verteilt. Daher ist die Anzahl $X := \sum_{i=1}^{100} X_i$ aller Rosinen im Kuchenstück des Kunden binomialverteilt $B_{100, \frac{1}{8}}$. Somit gilt

$$P(X \geq 7) = 1 - P(X \leq 6) = 1 - \sum_{i=0}^6 \binom{100}{i} \left(\frac{1}{8}\right)^i \left(1 - \frac{1}{8}\right)^{100-i} \approx 0.973.$$

Nunmehr modifizieren wir das Beispiel ein wenig.

Ein Bäcker liefert 500 derartige Kuchen an eine Lebensmittelkette. Er hat $n = 50000$ Rosinen für die Kuchenmasse verwendet und ein Tester nimmt ein Probestück. Mit welcher Wahrscheinlichkeit sind mindestens 7 Rosinen im Probestück?

Man hat nun $m = 4000$ Kuchenstücke und daher gilt für die Anzahl Y der Rosinen im Kuchenstück des Testers $Y \sim B_{50000, \frac{1}{4000}}$. Daraus folgt

$$P(Y \geq 7) = 1 - \sum_{i=0}^6 \binom{50000}{i} \left(\frac{1}{4000}\right)^i \left(1 - \frac{1}{4000}\right)^{50000-i}.$$

Klarerweise kann man die Binomialkoeffizienten nun nicht mehr direkt numerisch berechnen, aber die durchschnittliche Anzahl $\frac{n}{m}$ von Rosinen pro Kuchenstück beträgt, wie im obigen Fall 12.5, und daher liegt die Vermutung nahe, dass sich die beiden Wahrscheinlichkeiten $P(X \geq 7)$ und $P(Y \geq 7)$ nicht allzu sehr unterscheiden werden.

Der folgende Satz zeigt, dass die Vermutung richtig ist.

Satz 5.6. Ist $(p_n)_{n \in \mathbb{N}}$ eine Folge aus $[0, 1]$, so folgt aus $\lim_{n \rightarrow \infty} n p_n = \tau > 0$

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\tau^k}{k!} e^{-\tau} \quad \forall k \in \mathbb{N}_0. \quad (5.7)$$

Beweis. Aus $\lim_{n \rightarrow \infty} \frac{\binom{n}{k}}{n^k} = 1$ (siehe oben) und $\lim_{n \rightarrow \infty} n p_n = \tau$ folgt

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} &= \frac{1}{k!} \lim_{n \rightarrow \infty} \frac{\binom{n}{k}}{n^k} \lim_{n \rightarrow \infty} (n p_n)^k \lim_{n \rightarrow \infty} (1 - p_n)^{n-k} \\ &= \frac{\tau^k}{k!} \lim_{n \rightarrow \infty} (1 - p_n)^{n-k} = \frac{\tau^k}{k!} \lim_{n \rightarrow \infty} e^{(n-k) \ln(1-p_n)} = \frac{\tau^k}{k!} \lim_{n \rightarrow \infty} e^{-(n-k) p_n} \\ &= \frac{\tau^k}{k!} \lim_{n \rightarrow \infty} e^{-n p_n} \lim_{n \rightarrow \infty} e^{k p_n} = \frac{\tau^k}{k!} e^{-\tau} e^0 = \frac{\tau^k}{k!} e^{-\tau}. \end{aligned}$$

□

Wegen des obigen Satzes gilt $P(Y \leq 6) = \sum_{k=1}^6 \frac{(12.5)^k}{k!} e^{-12.5} \approx 0.03456739$, bzw. $P(Y \geq 7) = 1 - P(Y \leq 6) \approx 0.965$.

Definition 5.7. Eine Zufallsvariable X mit Wertebereich \mathbb{N}_0 und

$$P X^{-1}(k) = P(X = k) = \frac{\tau^k}{k!} e^{-\tau}, \quad k \in \mathbb{N}_0, \quad \tau > 0$$

wird poissonverteilt mit Parameter τ genannt (i.Z.: $X \sim P_\tau$).

Ist p_n die Wahrscheinlichkeit für das Auftreten eines Alternativereignisses (Rosinen im Kuchenstück), so ist $n p_n$ die durchschnittliche Anzahl dieser Ereignisse pro Einheit (Kuchenstück, Strecken-, Flächen-, Raum- oder Zeiteinheit). Die Voraussetzung des obigen Satzes besagt daher, dass die Binomialverteilung durch die Poissonverteilung approximiert werden kann, wenn die durchschnittliche Anzahl von Ereignissen pro Einheit annähernd konstant τ ist. Ändert man die Einheit auf das t -fache der ursprünglichen Einheit, so wird auch die durchschnittliche Anzahl von Ereignissen dementsprechend geändert und man erhält dafür $\tau_{neu} = t \tau_{alt}$. Teilt man die Kuchen beispielsweise in 10 Stücke an Stelle von 8, so gilt: $t = \frac{4}{5} \Rightarrow \tau_{neu} = \frac{4}{5} \cdot 12.5 = 10$, und die Anzahl von Rosinen in den kleineren Stücken wird etwa P_{10} verteilt sein.

Beispiel 5.8. In Österreich ereignen sich ca. 40000 Unfälle mit Personenschaden pro Jahr, das sind ca. 109.59 Unfälle pro Tag. X_d , die Anzahl von Unfällen pro Tag, ist daher annähernd $P_{109.59}$ verteilt, und X_h , die Anzahl von Unfällen pro Stunde, ist poissonverteilt mit $\tau \approx \frac{109.59}{24} \approx 4.57$, wenn sich die Unfälle gleichmäßig über das Jahr und auch gleichmäßig über den Tag verteilen.

Ob diese Annahme zutrifft, sollte man im obigen Beispiel kritisch hinterfragen, da die Unfallhäufigkeit im Sommer vermutlich anders als im Winter und am Tag anders als in der Nacht sein wird.

Definition 5.9. Falls für eine Funktion $\alpha(t)$ und eine Funktion $\beta(t)$ gilt:

$$\lim_{t \rightarrow 0} \frac{\alpha(t)}{\beta(t)} = 0, \quad (5.8)$$

so sagt man α ist ein klein-o von $\beta(t)$, (i.Z. $\alpha(t) = o(\beta(t))$).

Falls ein $M < \infty$ existiert mit

$$\limsup_{t \rightarrow 0} \left| \frac{\alpha(t)}{\beta(t)} \right| \leq M, \quad (5.9)$$

so nennt man $\alpha(t)$ ein Groß-O von $\beta(t)$ (i.Z. $\alpha(t) = O(\beta(t))$).

Ist $X_t, t > 0$ die Anzahl der Ereignisse in einer Einheit der Größe t (das kann eine Längen-, Flächen-, Raum- oder auch Zeiteinheit sein), so folgt aus $X_1 \sim P_\tau$, wie oben erwähnt, $X_t \sim P_{t\tau}$. Für kleines t gilt dann

$$P(X_t = 1) = \tau t e^{-\tau t} = \tau t \left(1 - \tau t + \frac{\tau^2 t^2}{2} - \dots \right) = \tau t + o(t) \quad (5.10)$$

$$\begin{aligned} P(X_t > 1) &= e^{-\tau t} \sum_{k=2}^{\infty} \frac{(\tau t)^k}{k!} = e^{-\tau t} (\tau t)^2 \sum_{j=0}^{\infty} \frac{(\tau t)^j}{(j+2)!} \\ &\leq e^{-\tau t} (\tau t)^2 \sum_{j=0}^{\infty} \frac{(\tau t)^j}{j!} = e^{-\tau t} (\tau t)^2 e^{\tau t} = \tau^2 t^2 = o(t). \end{aligned} \quad (5.11)$$

Wir kehren nochmals zu unserem Beispiel 5.5 mit n Rosinen und m Kuchenstücken zurück: Ist X_n die Anzahl der Rosinen in einem Stück und Y_n die Anzahl in einem 2-ten Stück, so gilt

$$\begin{aligned} P(X_n = i, Y_n = k) &= P(X_n = i) P(Y_n = k | X_n = i) \\ &= \binom{n}{i} \left(\frac{1}{m} \right)^i \left(1 - \frac{1}{m} \right)^{n-i} \binom{n-i}{k} \left(\frac{1}{m-1} \right)^k \left(1 - \frac{1}{m-1} \right)^{n-i-k} \end{aligned}$$

Mit $n \rightarrow \infty$ und $\lim_{n \rightarrow \infty} \frac{n}{m} = \tau$, konvergiert die obige Wahrscheinlichkeit gegen

$$\frac{\tau^i e^{-\tau}}{i!} \frac{\tau^k e^{-\tau}}{k!} = \lim_n P(X_n = i) P(Y_n = k).$$

X_n und Y_n sind also asymptotisch unabhängig.

Daher kann man annehmen, dass für die Anzahlen X_t und Y_t von Ereignissen in 2 disjunkten Einheiten der Größe t stets gilt

$$X_t \text{ und } Y_t \text{ sind unabhängig.} \quad (5.12)$$

Der folgende Satz zeigt umgekehrt, dass $X_1 \sim P_\tau$, wenn die Bedingungen (5.10), (5.11), und (5.12) erfüllt sind.

Satz 5.10. *Zerlegt man eine Einheit in disjunkte kleine Stücke der Größe t , bezeichnet man mit $X_{i,t}$ die Anzahl von Ereignissen im i -ten Stück, so folgt aus*

1. $P(X_{i,t} = 1) = \tau t + o(t) \quad \forall i$
2. $P(X_{i,t} > 1) = o(t) \quad \forall i$
3. $i \neq j \Rightarrow X_{i,t} \text{ und } X_{j,t} \text{ sind unabhängig,}$

dass die Anzahl X_1 von Ereignissen in einer Einheit der Größe 1 P_τ verteilt ist.

Beweis. Es gilt

$$P(X_1 = 0) = P\left(\bigcap_{i=1}^n [X_{i, \frac{1}{n}} = 0]\right) = \left(1 - \frac{\tau}{n} - o\left(\frac{1}{n}\right)\right)^n \rightarrow e^{-\tau}.$$

$$\begin{aligned} P(X_1 = 1) &= \sum_{i=1}^n P(X_{i, \frac{1}{n}} = 1) \prod_{j \neq i} P(X_{j, \frac{1}{n}} = 0) \\ &= n \frac{\tau}{n} \left(1 - \frac{\tau}{n}\right)^{n-1} + o\left(\frac{1}{n}\right) \rightarrow \tau e^{-\tau}. \end{aligned}$$

Ist $A_k := \bigcup_{1 \leq i_1 < \dots < i_k \leq n} [X_{i_j, \frac{1}{n}} = 1, \forall j = 1, \dots, k, X_{l, \frac{1}{n}} = 0, l \notin \{i_1, \dots, i_k\}]$,
so gilt $[X_1 = k] \supseteq A_k$ und daraus folgt

$$P(X_1 = k) \geq \binom{n}{k} \left(\frac{\tau}{n}\right)^k \left(1 - \frac{\tau}{n}\right)^{n-k} \rightarrow \frac{\tau^k e^{-\tau}}{k!}. \quad (5.13)$$

Umgekehrt gilt $[X_1 = k] \subseteq A_k \cup \bigcup_{i=1}^n [X_{i, \frac{1}{n}} > 1]$, und daraus folgt

$$P(X_1 = k) \leq \binom{n}{k} \left(\frac{\tau}{n}\right)^k \left(1 - \frac{\tau}{n}\right)^{n-k} + n o\left(\frac{1}{n}\right) \rightarrow \frac{\tau^k e^{-\tau}}{k!}. \quad (5.14)$$

Aus (5.13) und (5.14) folgt $P(X_1 = k) \rightarrow \frac{\tau^k e^{-\tau}}{k!}$.

□

5.5 Die geometrische Verteilung

Zu einer Folge (X_i) unabhängiger Zufallsvariabler mit $X_i \sim B_p$, $0 \leq p \leq 1$ soll die Verteilung von $Y := \min\{n \in \mathbb{N} : X_n = 1\}$ bestimmt werden.

Wegen $[Y > n] = \bigcap_{i=1}^n [X_i = 0]$ gilt $P(Y > n) = (1-p)^n$. Daraus folgt

$$P(Y = n) = P(Y > n-1) - P(Y > n) = (1-p)^{n-1} - (1-p)^n = p(1-p)^{n-1}.$$

Definition 5.11. Eine Zufallsvariable Y mit Wertebereich \mathbb{N} und

$$P(Y = n) = (1-p)^{n-1}p \quad \forall n \in \mathbb{N}$$

wird *geometrisch verteilt* genannt mit Parameter p (i.Z.: $Y \sim G_p$).

So ist etwa die Anzahl von Würfeln mit einem fairen Würfel, die man bis zur ersten „6“ benötigt, geometrisch verteilt mit $p = \frac{1}{6}$.

Bemerkung 5.12. Mitunter nennt man $\tilde{Y} := Y - 1$ die Anzahl der Versuche, die auf 0 enden (das ist die Anzahl der „Misserfolge“) geometrisch verteilt. Dann hat man als Wertebereich \mathbb{N}_0 und es gilt

$$P(\tilde{Y} = n) = (1-p)^n p \quad n \in \mathbb{N}_0.$$

Geometrisch verteilte Zufallsvariablen besitzen eine wichtige Eigenschaft.

Wegen $P(Y > n) = (1-p)^n \quad \forall n \in \mathbb{N}$ gilt

$$\begin{aligned} P(Y > n+m | Y > m) &= \frac{P([Y > n+m] \cap [Y > m])}{P(Y > m)} \\ &= \frac{P(Y > n+m)}{P(Y > m)} = \frac{(1-p)^{n+m}}{(1-p)^m} = (1-p)^n = P(Y > n). \end{aligned}$$

Dies bedeutet, dass die Anzahl der bis zu einem Zeitpunkt m durchgeführten (erfolglosen) Versuche keinen Einfluss auf die ab diesem Zeitpunkt bis zum ersten „Erfolg“ nötige Anzahl von Versuchen hat.

Man sagt daher, dass geometrisch verteilte Zufallsvariable *gedächtnislos* sind. Aber man kann sogar zeigen, dass unter allen Zufallsvariablen mit Wertebereich \mathbb{N} nur die geometrisch verteilten *gedächtnislos* sind.

Satz 5.13. Gilt für eine Zufallsvariable X mit Wertebereich \mathbb{N}

$$P(X > n+m | X > m) = P(X > n) \quad \forall n, m \in \mathbb{N},$$

so ist X geometrisch verteilt mit $p := P(X = 1)$.

Beweis. Mit $q := 1-p = P(X > 1)$ gilt

$$\begin{aligned} P(X > 2) &= P(X > 1)P(X > 2 | X > 1) \\ &= P(X > 1)P(X > 1) = P(X > 1)^2 = q^2. \end{aligned}$$

Angenommen es gilt $P(X > n - 1) = q^{n-1}$, so folgt daraus

$$\begin{aligned} P(X > n) &= P(X > n - 1) P(X > n | X > n - 1) \\ &= P(X > n - 1) P(X > 1) = q^n. \end{aligned}$$

Somit ist mit vollständiger Induktion $P(X > n) = q^n \quad \forall n \in \mathbb{N}$ gezeigt, und daraus folgt sofort $X \sim G_p$. \square

5.6 Die negative Binomialverteilung

Die *negative Binomialverteilung* oder *Pascalverteilung* ist eine Verallgemeinerung der geometrischen Verteilung. Sie ist die Verteilung der Anzahl Y von unabhängigen, alternativverteilten Versuchen mit fester *Erfolgswahrscheinlichkeit* p , die man benötigt bis r -mals eine „Eins“ auftritt. Bezeichnet man die Ausgänge der einzelnen Versuche mit $X_i \sim B_p$, so kann man Y formal anschreiben als $Y := \min \left\{ n : \sum_{i=1}^n X_i = r \right\}$.

Die r -te „Eins“ tritt gerade dann beim n -ten Versuch auf, wenn dieser Versuch auf *Eins* ausgeht und von den vorangegangenen $n - 1$ Versuchen $r - 1$ eine *Eins* ergeben und $n - 1 - (r - 1) = n - r$ eine *Null*. Mit den Bezeichnungen $I := \{i_1, \dots, i_{r-1}\}$ und $N_{n-1} := \{1, \dots, n - 1\}$ gilt daher

$$[Y = n] = \bigcup_{I \subseteq N_{n-1}} [X_j = 1, \forall j \in I, X_k = 0, \forall k \in N_{n-1} \setminus I, X_n = 1].$$

Da gilt $P(X_j = 1, \forall j \in I, X_k = 0, \forall k \in N_{n-1} \setminus I, X_n = 1) = p^r (1-p)^{n-r}$ für alle $I \subseteq N_{n-1}$ und weil es $\binom{n-1}{r-1}$ Möglichkeiten zur Auswahl von I aus N_{n-1} gibt, ergibt sich die Wahrscheinlichkeit von $[Y = n]$ zu

$$P(Y = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad \forall n \in \{r, r+1, \dots\}.$$

Definition 5.14. Eine Zufallsvariable Y mit Wertebereich $\{r, r+1, \dots\}$, $r \in \mathbb{N}$ nennt man *negativ binomialverteilt* mit den Parametern $r \in \mathbb{N}$ und $p \in [0, 1]$ (i.Z.: $Y \sim \text{neg}B_{r,p}$), wenn gilt

$$P(Y = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n \geq r. \quad (5.15)$$

Die geometrische Verteilung ist daher der Spezialfall einer negativen Binomialverteilung mit $r = 1$.

Bemerkung 5.15. Oft bezeichnet man $\tilde{Y} := Y - r$, also die Anzahl der Nullen vor der r -ten Eins als *negativ binomialverteilt*, dies vor allem deshalb, weil \tilde{Y} für jeden Parameter r den Wertebereich \mathbb{N}_0 besitzt. Für \tilde{Y} gilt

$$P(\tilde{Y} = n) = \binom{n+r-1}{r-1} p^r (1-p)^n, \quad n \in \mathbb{N}_0. \quad (5.16)$$

Bemerkung 5.16. Man kann für $x \in \mathbb{R}$ verallgemeinerte Binomialkoeffizienten definieren durch

$$\binom{x}{k} := \frac{x(x-1)\dots(x-k+1)}{k!}.$$

Damit gilt

$$\begin{aligned} \binom{n+r-1}{r-1} &= \binom{n+r-1}{n} = \frac{(n+r-1)\dots(r+1)r}{n!} \\ &= \frac{(-1)^n (-r)(-r-1)\dots(-r-n+2)(-r-n+1)}{n!} = (-1)^n \binom{-r}{n}. \end{aligned}$$

Setzt man dies in (5.16) ein, so erhält man

$$P(\tilde{Y} = n) = \binom{-r}{n} (-1)^n p^r (1-p)^n = \binom{-r}{n} p^r (p-1)^n,$$

woraus sich der Name „negative Binomialverteilung“ erklärt.

Beispiel 5.17. 2 Schachspieler A und B bestreiten ein Turnier. Derjenige der als erster 21 Einzelspiele für sich entscheiden kann, gewinnt das Turnier. Wie groß ist die Wahrscheinlichkeit, dass A das Turnier gewinnt, wenn A jedes Einzelspiel mit Wahrscheinlichkeit $p = 0.6$ gewinnt und wenn man annimmt, dass die Ausgänge der einzelnen Spiele unabhängig voneinander sind?

Lösung: \tilde{Y} die Anzahl der Niederlagen von Spieler A bis zum 21-ten Sieg ist $negB_{21,0.6}$ verteilt. Daher gilt

$$P(\tilde{Y} = n) = \binom{n+20}{20} p^{21} (1-p)^n.$$

und A gewinnt das Turnier, wenn $[\tilde{Y} \leq 20]$. Die Wahrscheinlichkeit, dass A das Turnier gewinnt, ergibt sich daher zu

$$P(\tilde{Y} \leq 20) = F_{\tilde{Y}}(20) \approx 0.90348.$$

Wir bezeichnen nun mit Z_i die Anzahl der Versuche bis zur i -ten Eins, d.h.

$$\begin{aligned} Z_1 &:= \min\{n : X_n = 1\}, \\ Z_2 &:= \min\{n > Z_1 : X_n = 1\}, \\ &\vdots \\ Z_r &:= \min\{n > Z_{r-1} : X_n = 1\}.. \end{aligned}$$

Dann gilt für die Zufallsvariablen $T_1 := Z_1$, $T_i := Z_i - Z_{i-1}$, $i > 1$

$$[T_i = n] = [X_{Z_{i-1}+1} = 0, \dots, X_{Z_{i-1}+n-1} = 0, X_{Z_{i-1}+n} = 1].$$

Das Ereignis $[T_i = n]$ ist daher unabhängig von den Versuchen $X_1, \dots, X_{Z_{i-1}}$ und daher auch unabhängig von T_1, \dots, T_{i-1} . Wegen $\tilde{Y} = Z_r = \sum_{i=1}^r T_i$ kann daher jede Zufallsvariable $Y \sim negB_{r,p}$ dargestellt werden als eine Summe von unabhängigen Zufallsvariablen $T_i \sim G_p$, $i = 1, \dots, r$.

