Wahrscheinlichkeitsrechnung
WS 94/95

gehalten von Univ.Doz. Erich Neuwirth
(Zusammenfassung der Vorlesung[1])

Bemerkung: Diese Unterlage will kein Ersatz des Besuchs der Vorlesung sein, sie dient nur als Erinnerungshilfe und Abgrenzung des behandelten Stoffes.

Grundzüge der Kombinatorik

Kombinatorik handelt vom Abzählen von Dingen (Zählen ohne wirklich zu zählen).
Es gibt eine Grundmenge (von n Dingen). Daraus werden k Dinge ausgewählt. Die Auswahl kann auf verschiedene Arten geschehen: mit Wiederholungen oder ohne Wiederholungen, und mit Berücksichtigung der Reihenfolge oder ohne Berücksichtigung der Reihenfolge. Auswahlen ohne Berücksichtigung der Reihenfolge nennt man auch Kombinationen, Auswahlen mit Berücksichtigung der Reihenfolge heißen Variationen.
Die allgemeine Formulierung des Problems läßt in der Grundmenge beliebige Elemente zu. Die Darstellung wird aber einfacher, wenn wir als Grundmenge des Umfangs n die Zahlen 1,2,...n wählen.

Kombinationen ohne Wiederholung

Repräsentation durch aufsteigende Folgen.
Beispiel: Jede Auswahl von 4 Zahlen aus 6, bei der Duplikate nicht erlaubt sind, und bei der die Reihenfolge keine Rolle spielt, kann als aufsteigende Folge von 4 Zahlen dargestellt werden. Das letzte Element dieser Folge kann höchstens 6 sein. 1234, 1246, 1346 ... sind Beispiele solcher Folgen.
Diese Folgen können wir danach unterscheiden, ob das letzte Element 6 ist oder nicht. Wenn wir diese beiden Teilmengen von Folgen abzählen, und die beiden Anzahlen addieren, haben wir die Gesamtzahl. Die Zahl der Folgen mit höchstens 5 an letzter Stelle ist genau die Zahl der 4 aus 5-Folgen. Andrerseits ist jede Folge mit 6 an letzter Stelle eine verlängerte 3 aus 5-Folge. Daher ist die Zahl der 4 aus 6-Folgen gleich der Summe der Zahl der 4 aus 5-Folgen und der Zahl der 3 aus 5-Folgen. Dieses Prinzip ist allgemein anwendbar und gilt für beliebige Folgen der Länge mindestens 2. Außerdem ist die Zahl der 1 aus x-Folgen für jedes x gleich diesem x und die Zahl der y aus 1 Folgen gleich 0 für y größer als 2.
Zusammenfassung als Tabelle:

	Plätze					
Objekte	1	2	3	4	5	6
1	1	0	0	0	0	0
2	2					
3	3					
4	4					
5	5					
6	6			+		

Die klassische algebraische Schreibweise dieser Beziehungen ist:

---

[1] Fassung vom 14. April 1995

$$f_1(n,1) = n$$
$$f_1(1,k) = 0 \text{ für alle } k > 1$$
$$f_1(n,k) = f_1(n-1,k) + f_1(n-1,k-1) \text{ sonst}$$

Wenn wir jetzt noch eine Spalte mit 0 Plätzen einführen, diese Spalte mit 1 füllen, und die Spalte mit der Nummer 1 ebenfalls durch die rekursive Summenbeziehung definieren, erhalten wir folgende Tabelle:

	Plätze					
Objekte	0	1	2	3	4	5
1	1	0	0	0	0	0
2	1					
3	1					
4	1					
5	1					
6	1			+		

erhalten wir folgende Beziehung:

$$f_1(n,0) = 1$$
$$f_1(1,k) = 0 \text{ für alle } k > 1$$
$$f_1(n,k) = f_1(n-1,k) + f_1(n-1,k-1) \text{ sonst}$$

Üblicherweise wird eine andere Bezeichnung verwendet:

$$f_1(n,k) = \binom{n}{k}$$

Da in dieser Tabelle jede Zahl zweimal in die Zeile darunter “sickert”, verdoppelt sich die Zeilensumme von Zeile zu Zeile, daher gilt

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Wir können auch folgende Überlegung anstellen:
In jede 3 aus 6-Folge können wir 3 Elemente „einfügen“, die in dieser Folge nicht verwendet wurden, um eine 4 aus 6-Folge zu erhalten.
134 kann erweitert werden zu 1234, 1345 und 1346. Also sollte es 3 mal soviele 4 aus 6-Folgen wie 3 aus 6-Folgen geben. Das stimmte aber nicht, weil 1234 auch durch Einfügen der Zahl 1 in die Folge 234 entstanden sein könnte. Die neu entstandenen Folgen werden mehrfach gezählt. Da jedes Element der neuen Folge das neu eingefügte sein kann, gilt allgemein folgende Rekursion:

$$f_1(n,k) = f_1(n,k-1)\frac{(n-(k-1))}{k}$$

und das führt zur üblichen Formel:

$$\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{1.2\ldots k} = \frac{n!}{k!(n-k)!}$$

Tabellendarstellung dieser Formel:

	Plätze				
Objekte	1	2	3	4	5
1	1				
2	2				
3	3				
4	4		*(-)/		
5	5				

Kombinationen mit Wiederholung:

Reihenfolge spielt keine Rolle, Duplikate erlaubt:
Repräsentiert durch nichtfallende Folgen.
Also kann beispielsweise jede 4 aus 6 Folge eine 4 aus 5 Folge oder eine um die Zahl 6 verlängerte 3 aus 6 Folge sein. Allgemeine Tabellendarstellung:

	Plätze					
Objekte	0	1	2	3	4	5
1	1	0	0	0	0	0
2	1					
3	1					
4	1					
5	1					
6	1			+		

Formeldarstellung:

$$f_2(n,0) = 1$$
$$f_2(1,k) = 1 \text{ für alle } k > 1$$
$$f_2(n,k) = f_2(n-1,k) + f_2(n,k-1) \text{ sonst}$$

Das führt zu folgender Beziehung zwischen den beiden Tabellendarstellungen:

	Plätze					
Obj	1	2	3	4	5	6
1	1	0	0	0	0	0
2	2					
3	3	+				
4	4					
5	5		+			
6	6					

	Plätze					
Obj	1	2	3	4	5	6
1	1		1	1	1	1
2	2	+				
3	3		+			
4	4					
5	5					
6	6					

Eine Scherungsoperation liefert:

					5	0
				4	0	
			3	0		
		2	0			
Obj	1	0				
1	1		1	1	1	1
2	2	+				
3	3		+			
4	4					
5	5					
6	6					

Obj	1	2	3	4	5	6
1	1		1	1	1	1
2	2	+				
3	3		+			
4	4					
5	5					
6	6					

und daher gilt

$$f_2(n,k) = f_1(n+k-1,k)$$

Variationen mit Wiederholung:

Wieviele Folgen vorgegebener Länge k, die nur die Zahlen $1,2...n$ enthalten, gibt es?

$$n^k$$

Da die Besetzungen an den einzelnen Stellen unabhängig voneinander stattfinden, werden die einzelnen Möglichkeiten miteinander multipliziert.

Variationen ohne Wiederholung

Für die Besetzung der ersten Stelle gibt es n Möglichkeiten, für die zweite Stelle $n-1$ Möglichkeiten usw. Daher insgesamt

$$n(n-1)..(n-k+1) = \binom{n}{k}k!$$

Daher können wir die Resultate zusammenfassen:

		Wiederholungen	
		ohne	mit
Reihenfolge	ohne	$\binom{n}{k}$	$\binom{n+k-1}{k}$
	mit	$n(n-1)...(n-k+1)$ $=\binom{n}{k}k!$	n^{k}

Eine weitere Gleichung:
Pfeilverfolgung in folgenden Diagramm:

	Plätze					
Objekte	0	1	2	3	4	5
1	1	0	0	0	0	0
2	1					
3	1					
4	1					
5	1					
6	1			+		

liefert folgende Tatsache:

	Plätze					
Objekte	0	1	2	3	4	5
1	1	0	0	0	0	0
2	1					
3	1					
4	1					
5	1					
6	1					

Die dunkle Zelle ist die Summe aller hellen Zellen. In Formeln, und mit Rückübersetzung in die Binomialkoeffizienten bedeutet das:

$$\sum_{i=0}^{k}\binom{n+i}{i} = \binom{n+k+1}{k}$$

Weiters wurden folgende kombinatorischen Probleme behandelt:

Sterlingzahlen zweiter Art:

Anzahl aller nichttrivialen Partitionen einer Menge mit n Elementen in k Teilmengen.
Rekursionsformel:

$$S_2(n,k) = kS_2(n-1,k) + S_2(n-1,k-1)$$

Sterlingzahlen erster Art:

Anzahl aller Permutationen von n Elementen mit einer Zyklendarstellung aus genau k Zyklen

$$S_1(n,k) = nS_1(n-1,k) + S_1(n-1,k-1)$$

Das einfachste kombinatorische Problem ist die Frage nach der Anzahl der Permutationen, der Anordnungen von n Zahlen in beliebiger Reihenfolge. Man kann dieses Problem auch als Frage nach den Variationen ohne Wiederholung, bei denen die Platzzahl und die Objektzahl gleich ist, deuten.

(Endliche) Wahrscheinlichkeitsrechnung

Grundlagen

Grundbegriff der Wahrscheinlichkeitsrechnung ist der Wahrscheinlichkeitsraum.
Das ist eine Menge X und eine Funktion p, die jedem x aus X eine Zahl $p(x)$ (genannt Wahrscheinlichkeit) zuordnet. Die x heißen Elementarereignisse. Die Summe aller Wahrscheinlichkeiten der Elementarereignisse muß 1 sein.
Beispiele:
$X = \{1,2,3,4,5,6\}$, $p(i) = \frac{1}{6}$ für $i = 1,2,\ldots 6$
beschreibt einen fairen Würfel
$X = \{k,a\}$, $p(a) = p(k) = \frac{1}{2}$
beschreibt den Wurf einer fairen Münze.
$X = \{k,a\}$, $p(a) = p(k) = p$
beschreibt den Wurf einer möglicherweise unfairen Münze.
In vielen Fällen ist nur X gegeben und aus sachlichen Gründen kann es sinnvoll scheinen, jedem einzelnen $x \in X$ dieselbe Wahrscheinlichkeit zuzuordnen. Viele klassische Aufgaben der Wahrscheinlichkeitsrechnung beginnen mit einem derartigen Modell der gleichwahrscheinlichen Elementarereignisse.
Der Wahrscheinlichkeitsbegriff wird dann auf beliebige Teilmengen A von X erweitert:

$$P(A) = \sum_{x \in A} p(x)$$

Wenn alle $p(x)$ gleich groß sind, dann kann man diese Definition auch noch folgendermaßen interpretieren:
Es gibt n Stück x, und A möge k davon enthalten. Man nennt dann die n Elemente auch die möglichen Fälle und die k, die zu A gehören, die günstigen Fälle. Daher gilt dann auch:

$$P(A) = \frac{k}{n} = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}$$

Viele Probleme beim Anwenden der Wahrscheinlichkeitsrechnung entstehen dadurch, daß begrifflich nicht sauber genug zwischen p und P unterschieden wird. p ist in der Regel von vornherein gegeben und muß nur eine einzige Bedingung erfüllen:

$$\sum_{x \in X} p(x) = 1$$

Aus dieser Eigenschaft lassen sich wichtige Eigenschaften von P ableiten:

$$P(X)=1$$
$$P(A\cup B)=P(A)+P(B) \quad \text{für} \quad A\cap B=\emptyset$$
$$P(\overline{A})=1-P(A)$$

Diese Eigenschaften trifft man manchmal auch unter dem Titel „Axiome von Kolmogoroff“ an. Im Falle unendlicher Wahrscheinlichkeitsräume muß man sie auch zur Grundlage der Wahrscheinlichkeitsrechnung machen. Im Falle der endlichen Wahrscheinlichkeitsräume reicht aber das Konzept der Funktion p (manchmal auch Wahrscheinlichkeitsmassenfunktion genannt) zur Schaffung sicherer Grundlagen aus.
Für alle praktischen Anwendungen dieses Modells kann man sagen, daß man Wahrscheinlichkeiten von zusammengesetzten Ereignissen (also Wahrscheinlichkeiten vom Typus $P(A)$) berechnen kann, indem man die passenden Wahrscheinlichkeiten von Elementarereignisse (vom Typus $p(x)$) addiert.

Anwendung: Inklusions-Exklusions-Prinzip

Typisches Beispiel: n Personen geben ihren Hut an der Garderobe ab. Dann werden die Garderobezettel zufällig vermischt. Wie hoch ist die Wahrscheinlichkeit, daß mindestens eine Person wieder ihren eigenen Hut erhält? Mathematische Übersetzung: Wie groß ist die Wahrscheinlichkeit, daß eine zufällig ausgewählte Permutation an einer Stelle i die Zahl i stehen hat. Man sagt dann auch, daß so eine Permutation einen Fixpunkt hat (Beispiel: die Permutation (3,2,4,1) hat den Fixpunkt 2, weil an der 2. Stelle die Zahl 2 steht).
Ausgangspunkt ist die Feststellung, daß leider folgende Formel im allgemeinen *nicht* gilt:

$$P(A\cup B)=P(A)+P(B)$$

Statt dessen gilt die Gleichung

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

Diese Gleichung läßt sich für Durchschnitte von mehr als 2 Mengen verallgemeinern.

$$P(\bigcup_{i=1}^{n}A_i)=\sum_{k=1}^{n}(-1)^{k-1}\sum_{i_1<i_2<\ldots i_k}P(\bigcap_{j=1}^{k}A_{i_j})$$

Diese Formel heißt auch Inklusions-Exklusions-Prinzip.
Anwendung auf unser Hutproblem:
A_i seien alle Permutationen mit Fixpunkt i . Dann gilt

$$P(\bigcap_{j=1}^{k}A_{i_j})=\frac{1}{n!}(n-k)(n-k-1)..1=\frac{(n-k)!}{n!}$$

und daher auch

$$\begin{aligned}P(\bigcup_{i=1}^{n}A_i)&=\sum_{k=1}^{n}(-1)^{k-1}\sum_{i_1<i_2<\ldots i_k}P(\bigcap_{j=1}^{k}A_{i_j})\\&=\sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}\frac{n!}{(n-k)!}\\&=\sum_{k=1}^{n}(-1)^{k-1}\frac{1}{k!}\cong 1-\frac{1}{e}\end{aligned}$$

Anwendungsbeispiel: Geburtstagsproblem

Wie groß ist die Wahrscheinlichkeit, daß in einer Gruppe von n Personen mindestens 2 am selben Tag Geburtstag haben?
Mathematisches Modell: Wir betrachten Folgen der Länge n gebildet aus Zahlen von 1 bis 365. Es gibt 365^n solche Folgen. Wieviele davon bestehen aus lauter verschiedenen Zahlen? Es gibt 365.364 solche Folgen der Länge 2, 365.364.363 solche Folgen der Länge 3. Allgemein gibt es $\prod_{i=1}^{n}(365-i+1)$ solche Folgen. Daher ist die gesuchte Wahrscheinlich-

keit $p_n = 1 - \frac{\prod_{i=1}^{n}(365-i+1)}{365^n}$. Wir können diese Formel auch so umschreiben: Sei q_n die Wahrscheinlichkeit, daß alle betrachteten Personen an verschiedenen Tagen Geburtstag haben, dann gilt

$$q_1 = 1$$
$$q_i = q_{i-1}\frac{365-i+1}{365}$$
$$p_i = 1 - q_i$$

Produktwahrscheinlichkeiten

Wichtig in der Wahrscheinlichkeitsrechnung ist, daß man ausgehend von bestimmten Wahrscheinlichkeitsräumen, die bestimmte Experimente beschreiben, durch Zusammensetzung weitere Wahrscheinlichkeitsräume konstruiert. Ein besonders wichtiges Modell ist dabei der Produktraum.
Beginnen wir zunächst mit dem Produktraum von zwei (möglicherweise verschiedenen) Wahrscheinlichkeitsräumen:
X mit Wahrscheinlichkeit p
Y mit Wahrscheinlichkeit q
Dann definieren wir einen neuen Wahrscheinlichkeitsraum mit der „Trägermenge“ $X \times Y$ und der Wahrscheinlichkeitsfunktion
$\overline{p}((x,y)) = p(x)q(y)$
Natürlich gilt $\sum_{x\in X, y\in Y} \overline{p}((x,y)) = \sum_{(x,y)\in X\times Y} \overline{p}((x,y)) = 1$
Diese Wahrscheinlichkeitsfunktion für Elementarereignisse läßt sich in gewohnter Weise zu einer Wahrscheinlichkeitsfunktion für zusammengesetzte Ereignisse erwei tern.
Außerdem gilt folgende wichtige Gleichung:
$\overline{P}(A\times B) = P(A)Q(B)$
Der Produktraum modelliert das simultane oder sequentielle Ausführen mehrerer Experimente, wobei der Ausgang jedes einzelnen Teilexperiments keinen Einfluß auf die Ausgänge der anderen Teilexperimente hat.
Der Produktraum mit lauter gleichen Komponenten modelliert die Mehrfachhintereinanderausführung eines Experiments, also beispielsweise das 10-malige Werfen einer Münze.
Wenn X der Grundraum ist, der zum einfachen Experiment gehört, dann bildet man das n-fache kartesische Produkt

$$\underbrace{X\times X\times\ldots\times X}_{n\text{-fach}}$$

Die Elemente des Produktraums sind endliche Folgen der Form $(x_1, x_{2,}\ldots, x_n)$, und wir können die Elementarwahrscheinlichkeiten für diese Elemente folgendermaßen definieren:

$$\overline{p}((x_1, x_{2,}\ldots, x_n)) = p(x_1)p(x_2)\ldots p(x_n) = \prod_{i=1}^{n} p(x_i)$$

(Hinter dieser Definition steckt die Verallgemeinerung des Prinzips, daß im Falle glei cher Elementarwahrscheinlichkeiten auf dem einfachen Grundraum die Multiplikation der Elementarwahrscheinlichkeiten der Komponenten wieder gleiche Elementarwahrscheinlichkeiten auf dem Produktraum liefert).
Beginnend mit dieser Elementarwahrscheinlichkeit auf dem Produktraum erhalten wir eine Wahrscheinlichkeit (für Mengen, also zusammengesetzte Ereignisse) auf dem Produktraum, indem wir die Wahrscheinlichkeiten der einzelnen Elementarereignisse in der Menge addieren. Für diese Wahrscheinlichkeit gilt folgendes Ergebnis:

$$\overline{P}(X\times X\ldots\times A\times\ldots\times B\times X\times\ldots\times X) = P(A)P(B)$$

In allen diesen Formeln sind p, P die Wahrscheinlichkeiten auf dem einfachen Grundraum und $\overline{p}, \overline{P}$ die entsprechenden Wahrscheinlichkeiten auf dem Produktraum.

Diese Beschreibung konstruiert den Produktraum aus mehreren identischen Faktoren. Natürlich kann man die Konstruktion auch analog für das Produkt zweier oder mehrerer verschiedenen Faktoren durchführen. Die meisten interessanten Anwendungen dieses Konstruktionsprinzips beziehen sich jedoch auf das Produkt von mehreren identischen Faktoren.

Anwendungsbeispiel: Hypergeometrische Verteilung

Es werden „Stichproben" fixen Umfangs aus einer Grundmenge vorgegebenen Umfangs gezogen.
Stichproben sind Teilmengen.
Wir gehen von einer Grundmenge mit N Elementen aus, davon haben M eine bestimmte Eigenschaft (Beispiel: Grundmenge insgesamt N Kugeln, davon M weiß und $N-M$ schwarz). Wir wählen zufällig eine Teilmenge vom Umfang n. Mit welcher Wahrscheinlichkeit enthält diese Teilmenge m weiße Kugeln?
Grundannahme ist, daß jede mögliche Teilmenge gleich wahrscheinlich ist. Es gibt insgesamt $\binom{N}{n}$ mögliche Teilmengen. Die Frage ist, wieviele davon genau m Elemente der M ausgezeichneten Elemente enthalten. Es gibt genau $\binom{M}{m}$ verschiedene Auswahlen der m Elemente und $\binom{N-M}{n-m}$ Auswahlen der restlichen Elemente. Insgesamt gibt es daher $\binom{M}{m}\binom{N-M}{n-m}$ Teilmengen des Umfangs n mit genau m der „interessanten" Elemente. Die gesuchte Wahrscheinlichkeit beträgt daher:

$$\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$

Bei diesem Modell der hypergeometrischen Verteilung ist zu beachten, daß die Elemente nicht der Reihe nach, sondern simultan gezogen werden („eine Handvoll auf einmal").

Unabhängigkeit und bedingte Wahrscheinlichkeit

Unter der bedingten Wahrscheinlichkeit von A gegeben B versteht man

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

(Die Ausgangsidee ist, den Anteil einer Unteruntergruppe – also beispielsweise der rauchenden Männer – an einer Untergruppe – der Männer – in der Grundgesamtheit – Männer und Frauen zusammen – zu berechnen. Wahrscheinlichkeiten sind ihrer Idee nach idealisierte Häufigkeiten!)
Selbstverständlich ist diese Definition nur sinnvoll, falls $P(B) > 0$.
Zwei Ereignisse A und B heißen unabhängig, falls gilt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

(In unserem Raucherbeispiel würde das bedeute, daß der Raucheranteil in der Gruppe der Männer gleich hoch ist wie der Raucheranteil in der Gesamtgruppe Männer und Frauen)
Falls der zugrundeliegende Wahrscheinlichkeitsraum X ein Produktraum und die vorliegende Wahrscheinlichkeit eine Produktwahrscheinlichkeit ist, kann man zeigen, daß bestimmte Ereignis unabhängig sind.

Genauer: falls

$$X = X_1 \times X_2$$
$$p(x_1, x_2) = p_1(x_1)p_2(x_2)$$

dann gilt für beliebige $A \subset X_1$ und $B \subset X_2$

$$P(A \times B) = P_1(A)P_2(B)$$

und somit

$$P(A \times X_2 | X_1 \times B) = \frac{P((A \times X_2) \cap (X_1 \times B))}{P(X_1 \times B)}$$
$$= \frac{P(A \times B)}{P_2(B)} = \frac{P_1(A)P_2(B)}{P_2(B)} = P_1(A)$$

Identifiziert man noch in etwas ungenauer Sprech- und Schreibweise Ereignisse $A \subset X_1$ mit $A \times X_2 \subset X_1 \times X_2$ (und analoges für $B \subset X_2$), dann kann man *ungenauer* auch schreiben:

$$P(A \cap B) = P_1(A)P_2(B) = P(A)P(B)$$

Ereignisse, die „in verschiedenen Komponenten" eines Produktraums liegen, sind also unabhängig, wenn die Gesamtwahrscheinlichkeit eine Produktwahrscheinlichkeit ist.
Für bedingte Wahrscheinlichkeiten gelten 2 wichtige Sätze:

Satz von der totalen Wahrscheinlichkeit

Wenn die Mengen A_i, $i = 1..n$ eine Partition der Grundmenge X sind, also $A_i \cap A_j = \emptyset$ für $i \neq j$ und $\bigcup_{i=1}^{n} A_i = X$, dann gilt

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Satz von Bayes

Wenn die Mengen A_i, $i = 1..n$ ein Partition der Grundmenge X sind, also $A_i \cap A_j = \emptyset$ für $i \neq j$ und $\bigcup_{i=1}^{n} A_i = X$, dann gilt

$$P(A_j | B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Wahrscheinlichkeitskerne und bedingte Wahrscheinlichkeiten.

Oft versucht man ein mehrstufiges Phänomen zu modellieren. Dabei nimmt man an, daß es für die erste Versuchsstufe eine Wahrscheinlichkeit gibt, die man kennt. Außerdem nimmt man an, daß man für die zweite Stufe eine Wahrscheinlichkeit kennt, wenn das Ergebnis der ersten Stufe vollständig bekannt ist. Genauer gesagt nehmen wir an, daß wir auf dem Raum X_1 eine Wahrscheinlichkeit p_1 haben und für jedes einzelne Elementarereignis $x_1 \in X_1$ eine Wahrscheinlichkeit $p_2(x_1, x_2)$ auf X_2 kennen. Die Funktion p_2 ist für jedes einzelne x_1 eine Wahrscheinlichkeit, d.h. es gilt $\sum_{x_2 \in X_2} p_2(x_1, x_2) = 1$ für alle x_1.

Wir können daher auch für jedes einzelne x_1 die zugehörige „Mengenwahrscheinlichkeit" $P_2(x_1, B)$ für $B \subset X_2$ betrachten. Eine Verallgemeinerung von p_2 auf Mengen als Argumente an Stelle der Elemente x_1 ist aber *nicht* sinnvoll.

Die Wahrscheinlichkeiten p_2 bzw. P_2 nennt man auch Wahrscheinlichkeitskern.
Wir können dann eine Wahrscheinlichkeit P auf $X_1 \times X_2$ folgendermaßen definieren:

$$p(x_1, x_2) = p_1(x_1) p_2(x_1, x_2)$$

Für die entsprechende (Mengen-)Wahrscheinlichkeit

$$P(A) = \sum_{(x_1, x_2) \in X_1 \times X_2} p_1(x_1) p_2(x_1, x_2)$$

gilt

$$P(\{x\} \times B) = p_1(x) P_2(x, B) = P_1(\{x\}) P_2(x, B)$$

und daher gilt auch

$$\begin{aligned} P(X_1 \times B | \{x\} \times X_2) &= \frac{P((X_1 \times B) \cap (\{x\} \times X_2))}{P(\{x\} \times X_2)} \\ &= \frac{P(\{x\} \times B)}{P(\{x\} \times X_2)} = \frac{p_1(x) P_2(x, B)}{p_1(x)} \\ &= P_2(x, B) \end{aligned}$$

Die bedingte Wahrscheinlichkeit die wir als Ergebnis eines bestimmten Typs von kombinierten Ereignissen erhalten, sind also die Wahrscheinlichkeitskerne, mit denen die Modellbildung beginnt. Wahrscheinlichkeitskerne sind in dieser Terminologie aber nur für bedingte Wahrscheinlichkeiten definiert, bei denen die Bedingung ein Elementarereignis, also eine einelementige Menge ist. Wahrscheinlichkeitskerne mit Mengen als Argumente „an der ersten Stelle“ sind nicht sinnvoll zu definieren.

Anwendungbeispiel: Ziehen ohne Zurücklegen

Wir haben eine Urne mit Kuglen mit den Zahlen 1,2... n. Die Wahrscheinlichkeit, eine dieser Zahlen zu ziehen ist $\frac{1}{n}$. Wir haben als

$$p_1(i) = \frac{1}{n} \text{ für } i = 1,2,..n$$

Bei der zweiten Ziehung ist die erste Kugel nicht mehr vorhanden (ohne Zurücklegen!). Daher gilt

$$p_2(i, j) = \frac{1}{n-1} \text{ für } j \neq i \text{ und } p_2(i, i) = 0 \text{ für alle } i$$

Für die Gesamtwahrscheinlichkeit P gilt dann

$$p(i, j) = P(\{(i, j)\}) = \frac{1}{n(n-1)} \text{ für } i \neq j \text{ und } p(i, i) = P(\{(i.i)\}) = 0$$

Zufallsvariable

Wir gehen von einem Wahrscheinlichkeitsraum X mit einer Wahrscheinlichkeit p (für Elementarereignisse) bzw. P (für zusammengesetzte Ereignisse) aus. Auf der Menge X ist noch eine Abbildung f, die in eine weitere Menge Y führt, definiert.
Dann kann man auf Y eine durch f induzierte (Elementar-)Wahrscheinlichkeit definieren:

$$\begin{aligned} &f: X \to Y \\ &p_f(y) = \sum_{x \in X, f(x) = y} p(x) = P(f^{-1}(\{y\})) \end{aligned}$$

Dabei ist f^{-1} nicht die inverse Funktion, sondern der Urbildoperator.

Es gibt auch ein durch f induziertes P_f. Man kann dieses P_f zunächst wie üblich über p_f definieren. Es stellt sich dann heraus, daß auch gilt:

$$P_f(A) = \sum_{x\in X, f(x)\in A} p(x) = P(f^{-1}(A))$$

Damit vermittelt jede Abbildung auch neue Wahrscheinlichkeiten. Alle Berechnungen sind sehr einfach, wenn die Abbildung f injektiv ist. Interessante Anwendungen entstehen in jenen Fällen, wo f nicht injektiv ist.
Diese Konstruktion kann für beliebige Funktionen durchgeführt werden. Falls der Werteraum der Abbildung f die reellen Zahlen sind, nennt man die Abbildung Zufallsvariable.

Beispiel: Binomialverteilung:

Wir beginnen mit dem Grundraum
$X = \{0,1\}$ und der Wahrscheinlichkeit $p(1) = p,\ p(0) = 1-p$
(Diese Notation ist nicht ganz konsistent, aber üblich, und man gerät dabei auch nicht in allzu große Schwierigkeiten.)
Als nächstes bilden wir den n-fachen Produktraum

$$X^n = \{\vec{x} = (x_1, x_2, \ldots, x_n), x_i \in X\}$$

$$p(\vec{x}) = \prod_{i=1}^{n} p(x_i)$$

Also hat eine Folge, die aus k Einsen und $(n-k)$ Nullen besteht, die Wahrscheinlichkeit $p^k(1-p)^{n-k}$
Nun betrachten wir die Abbildung

$$f: X^n \to R$$

$$f(\vec{x}) = \sum_{i=1}^{n} x_i \quad \text{für} \quad \vec{x} = (x_1, x_2, \ldots. x_n)$$

Die Folgen $\vec{x}$ bestehen nur aus 0 und 1, und die Abbildung f bildet einfach die Summe aller dieser Zahlen, sie zählt also die Anzahl der Einser in der Folge.
Uns interessiert jetzt die Wahrscheinlichkeit $p_f(k)$
Da alle Folgen mit derselben Anzahl Einser dieselbe Wahrscheinlichkeit haben, (nämlich $p^k(1-p)^{n-k}$ bei k Einsern), brauchen wir nur die Anzahl der Folgen mit genau k Einsern abzuzählen. Es gibt $\binom{n}{k}$ solche Folgen. Daher gilt für die gesuchte Wahrscheinlichkeit

$$P(f(\vec{x}) = k) = P(\vec{x} \in X^n : f(\vec{x}) = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Bei dieser Formel kann man vereinfachend davon sprechen, daß der Binomialkoeffizient eine Folge der Nichtinjektivität der Abbildung f ist, und daß der Ausdruck $p^k(1-p)^{n-k}$ deswegen entsteht, weil der „Basisraum" der Abbildung f ein Produktraum ist.

Beispiel: Poissonverteilung

Eine weitere wichtige Verteilung ist die Poissonverteilung. Diese Verteilung hat positive Wahrscheinlichkeitswerte für alle natürlichen Zahlen. Die Formel für die Elementarwahrscheinlichkeiten dieser Verteilung lauten:

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Sie hat folgende wichtige Eigenschaft:
Für eine Folge von Binomialverteilungen mit $n \to \infty$ und $p \to 0$ und $np \to \lambda$ konvergieren die Wahrscheinlichkeiten

$$\binom{n}{k} p^k (1-p)^{n-k} \to e^{-\lambda} \frac{\lambda^k}{k!}$$

Wir können also bei großem Stichprobenumfang n und kleiner Basiswahrscheinlichkeit p die Wahrscheinlichkeiten für eine Binomialverteilung näherungsweise durch die entsprechenden Werte einer Poissonverteilung ersetzen. In der Regel ist die numerische Berechnung der Poisson-Wahrscheinlichkeiten einfacher.

Beispiel: Binomialverteilung und hypergeometrische Verteilung

Es gilt auch folgende näherungsweise Beziehung:
Wenn für eine Folge hypergeometrischer Verteilung $M \to \infty$, $N \to \infty$ und $\frac{M}{N} \to p$, dann geht (für festes n)

$$p(k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \to \binom{n}{k} p^k (1-p)^{n-k}$$

Praktisch bedeutet das folgendes: wenn man ein Stichprobe vom Umfang n aus einer Grundgesamtheit vom Umfang N zieht und n sowohl im Vergleich zu M als auch im Vergleich zu $N-M$ (und daher auch im Vergleich zu N) klein ist, dann kann man die Berechnungsformel für die hypergeometrische Verteilung durch die Formel für die Binomialverteilung ersetzen und begeht erhält auf diese Weise einen brauchbaren Näherungswert. Die Berechnung nach der Binomial-Formel ist in der Regel wesentlich einfacher als nach der Formel für die hypergeometrische Verteilung.
Eine vereinfachnde Interpretation lautet: Ziehen aus sehr großen grundgesamtheiten ist praktisch dasselbe wie Wiederholung eines immer gleichen Experiments.

Erwartungswert und Varianz von Zufallsvariablen

Gegeben sei eine Zufallsvariable

$$f: X \to R$$

Dann ist der Erwartungswert dieser Zufallsvariablen folgendermaßen definiert:

$$E(f) = \sum_{x \in X} f(x) p(x)$$

In dieser Form der Definition werden die Abbildung f und die Wahrscheinlichkeit auf X explizit verwendet. Einfache Überlegungen zeigen aber, daß auch gilt:

$$E(f) = \sum_{r \in f(X)} r p(r)$$

Es reicht also, die von f induzierte Wahrscheinlichkeit auf R zu kennen. Man muß X und die Wahrscheinlichkeiten darauf gar nicht kennen, um den Erwartungswert einer auf X definierten Zufallsvariablen berechnen zu können.
Mit Hilfe dieser Überlegungen läßt sich der Erwartungswert einer binomialverteilten Zufallsvariablen berechnen, und man erhält den Wert np.
Eine weitere wichtige Größe ist die Varianz einer Zufallsvariablen, die folgendermaßen definiert ist:

$$V(F) = \sum_{x \in X} p(x)(f(x) - E(f))^2 = \sum_{r \in f(X)} p_f(r)(r - E(f))^2$$

Ein weiterer wichtiger Begriff ist die Standardabweichung einer Zufallsvariablen:

$$\sigma(f)=\sqrt{V(f)}$$

Wie wir später sehen werden, erweist es sich aus mathematischen Gründen oft als sinnvoll, aus einer Zufallsvariablen f eine weitere Zufallsvariable g abzuleiten:

$$g=\frac{f-E(f)}{\sigma(f)}$$

g heißt dann standardisierte Zufallsvariable.

g und f haben normalerweise natürlich nicht dieselbe Wahrscheinlichkeitsverteilung.

Faltung, Summe unabhängiger Zufallsvariabler

In vielen Anwendungsfällen tritt eine Situation auf, die man folgendermaßen modellieren kann:

Gegeben sind zwei Zufallsvariable f und g, die als unabhängig vorausgesetzt werden. Es gilt also:

$$P(f=x,g=y)=P(f=x)P(g=y)=p_f(x)p_g(y)$$

Wir können dann die Wahrscheinlichkeiten der Verteilung der Summe von f und g folgendermaßen ausrechnen:

$$P(f+g=z)=\sum_y P(f=z-y)P(g=y)=\sum_y p_f(z-y)p_g(y)$$

Falls die bei den Zufallsvariablen auftretenden Werte (also die möglichen x- und y-Werte) alle ganzzahlig sind, läßt sich diese Formel auch als inneres Produkt zweier Wahrscheinlichkeitsvektoren interpretieren. Dabei muß allerdings der eine Vektor „verschoben und auf den Kopf gestellt werden".

Ein Beispiel, in dem dieses Modell anzuwenden ist, wäre das Würfeln mit 2 Würfeln mit Summation der beiden Augenzahlen. Dabei könnte man sowohl von zwei verschiedenen Würfeln als auch von einem Würfel, der zweimal geworfen wird, ausgehen.

Diese Konstruktion kann auch verallgemeinert werden, wir können auch Summen von mehr als zwei unabhängigen Zufallsvariablen betrachten. Die Wahrscheinlichkeitsverteilung einer solchen Summe läßt sich dann einfach mit Hilfe rekursiver Formeln berechnen. Anwendungsbeispiele für dieses Modell sind Glücksspiele, bei denen oft hintereinander immer wieder dasselbe Spiel gespielt wird. (also etwa einen Abend lang Roulette, wobei man immer auf die selbe Art von Chancen setzt).

Grenzwertsätze

Summen „vieler" Zufallsvariabler haben einige zunächst überraschende Eigenschaften:

Im folgenden seine $f_1, f_2, \ldots$ unabhängige Zufallsvariable, die alle dieselbe Verteilung und daher auch den selben Erwartungswert $E(f_1)$ und dieselbe Standardabweichung $\sigma(f_1)$ besitzen. Dann gilt:

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n f_i}{n}-E(f_1)\right|\geq\varepsilon\right)=0$$

Dieses Gesetz heißt Gesetz der großen Zahlen. Es besagt im wesentlichen, daß bei großer Anzahl von Versuchswiederholungen das arithmetische Mittel der Versuchsausgänge gegen den Erwartungswert eines einzelnen Experiments konvergiert.

Zum Beweis dieses Gesetzes verwendet man die Tschebyscheffsche Ungleichung:

$$P(|f-E(f)|\geq t)\leq \tfrac{1}{t^2}V(f)$$

Der zweite wichtige Grenzwertsatz ist der Zentrale Grenzwertsatz:

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} f_i - nE(f_1)}{\sigma\sqrt{n}} \le x\right) = \Phi(x)$$

Dabei ist $\Phi(x)$ die Verteilungsfunktion der Standardnormalverteilung, und sie ist definiert durch folgende Gleichung:

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{z^2}{2}}\, dz$$

Da viele Anwendungsbeispiele einigermaßen realistisch als Summe unabhängiger identisch verteilter zufälliger Variabler gedeutet werden können, ist es mit Hilfe dieses Satzes oft möglich, näherungsweise die Wahrscheinlichkeit interessierender Ereignisse auszurechnen.

Beispiel: Wie groß ist die Wahrscheinlichkeit, beim Würfeln mit 100 Würfeln eine Augensumme zwischen 325 und 375 zu würfeln?
Unsere Zufallsvariablen $f_1, f_2, \ldots$ nehmen alle die Werte 1 bis 6 mit Wahrscheinlichkeit je 1/6 an. Daher gilt

$$E(f_1) = \tfrac{7}{2} = 3.5$$

$$\sigma(f_1) = \sqrt{\tfrac{35}{12}} = 1.708$$

und daher gilt auch

$$P(325 \le \sum_{i=1}^{100} f_i \le 375) = P(\frac{325 - 100\times 3.5}{10\sqrt{\frac{35}{12}}} \le \frac{\sum_{i=1}^{100} f_i - nE(f_1)}{\sigma\sqrt{n}} \le \frac{375 - 100\times 3.5}{10\sqrt{\frac{35}{12}}})$$

Der mittlere Ausdruck in dieser Ungleichungskette ist jetzt in der Form, die dem zentralen Grenzwertsatz entspricht. Daher gilt, daß die gesuchte Wahrscheinlichkeit sich folgendermaßen berechnen läßt:

$$\Phi\left(\frac{375 - 100\times 3.5}{10\sqrt{\frac{35}{12}}}\right) - \Phi\left(\frac{325 - 100\times 3.5}{10\sqrt{\frac{35}{12}}}\right) = 0.928 - 0.072 = 0.856$$