Statistik für Informatiker, SS 2019

1 Grundlagen aus der Wahrscheinlichkeitstheorie

1.1 Ereignisse, Zufallsvariablen und Wahrscheinlichkeiten

Matthias Birkner

http://www.staff.uni-mainz.de/birkner/StatInfo19/

29.4.2019

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Worum geht es?

In vielen Situationen tritt „Zufall“ oder „Ungewissheit“ auf, in diesem Kapitel geht es uns darum, den üblichen mathematischen Rahmen kennen zu lernen, um solche Phänomene zu beschreiben.

Beispiele.

- Im Auftakt-Beispiel in Kapitel 0 haben wir ein zufälliges Pixel Z aus (einer diskretisierten Version von) $[0,1]^2$ gewählt.
- „Glücksspiel-Situationen“: Wir werfen eine Münze oder einen Würfel, spielen Lotto, schauen ein gut gemischtes Pokerblatt an, ...
- Wir befragen einen zufällig ausgewählten Studenten in der Mensa nach seinem Studiengang und danach, wieviel er heute in der Mensa ausgegeben hat.
- Wir übergeben n Zahlen in rein zufälliger Reihenfolge an einen Sortieralgorithmus;
  wir überprüfen zwei Polynome $f(x)$ und $g(x)$ vom Grad $d \gg 1$ auf Gleichheit, indem wir an k zufällig ausgewählten Stellen auf Gleichheit testen;
  wir betrachten einen Router, der Datenpakete an einen zufällig ausgewählten Nachbarknoten im Netzwerk schickt; ...

Mathematische Modellierung des Zufalls („Zutaten"):

Sei Ω (nicht-leere) Menge („Ergebnisraum" oder „Stichprobenraum" genannt), und sei $\mathcal{F} \subset 2^{\Omega}$ $(:= \{B : B \subset \Omega\})$.

$\omega \in \Omega$ heißt ein „Elementarereignis",

$A \in \mathcal{F}$ nennen wir ein „Ereignis",

insbesondere Ω ... „sicheres Ereignis",
$\varnothing$ ... „unmögliches Ereignis"

Vorstellung/Interpretation:

„der Zufall" wählt ein $\omega \in \Omega$, wir sagen „A tritt ein", wenn $\omega \in A$.

Operationen

(Mengenoperationen und ihre Interpretation für Ereignisse)

Für $A, B \in \mathcal{F}$

$A^c := \Omega \setminus A$	…	„A tritt nicht ein" (A^c heißt Gegen- oder Komplementärereignis von A)
$A \cap B$	…	„A und B treten ein"
$A \cup B$	…	„A oder B treten ein"
$A \subset B$	…	„A impliziert B" (falls A eintritt, so auch B)

A und B heißen *disjunkt*, wenn $A \cap B = \emptyset$.

Wir notieren auch $A \setminus B := A \cap B^c = \{\omega : \omega \in A, \omega \notin B\}$ („A tritt ein, aber B nicht").

Forderungen für Ereignisse

Wir fordern, dass $\mathcal{F}$ erfüllt

$$\begin{aligned} &i) \quad \varnothing \in \mathcal{F}, \\ &ii) \quad A \in \mathcal{F} \;\Rightarrow\; A^c \in \mathcal{F}, \\ &iii) \quad A_1, A_2, \dots \in \mathcal{F} \;\Rightarrow\; \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}. \end{aligned} \tag{1}$$

$\mathcal{F} \subset 2^{\Omega}$, das (1), *i)–iii)* genügt, heißt eine σ*-Algebra* (über Ω),

die de Morgansche Regeln ergeben dann auch $\Omega = \varnothing^c \in \mathcal{F}$, $\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{F}$.

Forderungen für Wahrscheinlichkeiten

Eine Abbildung $P : \mathcal{F} \to [0, 1]$ mit

(N) $P(\Omega) = 1$ („Normierung") und

(A) $A_1, A_2, \dots \in \mathcal{F}$ paarweise disjunkt (2)

$$\Rightarrow P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n) \qquad \text{(„}\sigma\text{-Additivität")}$$

heißt ein *Wahrscheinlichkeitsmaß* (auf $(\Omega, \mathcal{F})$).

$P(A)$ heißt / nennen wir die *Wahrscheinlichkeit des Ereignisses A*.

Definition 1.1

Ein Tripel $(\Omega, \mathcal{F}, P)$, wobei $\mathcal{F} \subset 2^\Omega$ die Forderungen (1), *i)–iii)* und P die Forderungen (2), (N) und (A) erfüllt, heißt ein *Wahrscheinlichkeitsraum*.

Beispiel und Definition 1.2 (Diskreter Wahrscheinlichkeitsraum)

Ω endlich oder abzählbar, $p : \Omega \to [0, 1]$ Abbildung mit $\sum_{\omega \in \Omega} p(\omega) = 1$.

$$P(A) := \sum_{\omega \in A} p(\omega), \quad A \subset \Omega$$

definiert ein Wahrscheinlichkeitsmaß auf $(\Omega, 2^{\Omega})$.

p heißt die (Wahrscheinlichkeits-)*Gewichtsfunktion* von P, $p(\omega)$ heißt das (Wahrscheinlichkeits-)*Gewicht* von ω.

Dies erfüllt „offensichtlich“ Forderung (2).

Beispiele 1.3

1. (uniforme Wahl aus endlicher Menge, „Laplace-Experimente")
   $\#\Omega < \infty$ und $p(\omega) = \frac{1}{\#\Omega}$ in Bsp. 1.2, z.B.
    1. (Wurf eines fairen 6er-Würfels) $\Omega = \{1, 2, 3, 4, 5, 6\}$, $p(\omega) = 1/6$ für $\omega \in \Omega$
    2. (dreimaliger Wurf eines fairen 6er-Würfels) $\Omega = \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{1, 2, 3, 4, 5, 6\}$ für $i = 1, 2, 3\}$, $p((\omega_1, \omega_2, \omega_3)) = 1/6^3 = 1/216$
    3. (Auftakt-Beispiel aus Kapitel 0 mit Diskretisierung auf 32 Bit Genauigkeit) $\Omega = \{(x, y) : x = k/2^{32}, y = \ell/2^{32}$ mit $k, \ell \in \{0, 1, 2, \dots, 2^{32} - 1\}\}$, $p((x, y)) = 1/(2^{32})^2 = 2^{-64}$
    4. (n verschiedene Eingaben in zufälliger Reihenfolge) $\Omega = \{(x_1, x_2, \dots, x_n) : x_1, \dots, x_n \in \{1, 2, \dots, n\}$ paarweise verschieden$\}$ („symmetrische Gruppe der Ordnung n"), $p((x_1, x_2, \dots, x_n)) = 1/n!$

Beispiele 1.3 (Fortsetzung)

2. (ein verfälschter Münzwurf) $\Omega = \{\mathrm{Kopf}, \mathrm{Zahl}\}$, $p(\mathrm{Kopf}) = 0.6 = 1 - p(\mathrm{Zahl})$
3. (eine winziges Modell für Spam, das Sprache und Status einer Email betrachtet)

$$\Omega = \{\texttt{Deutsch}, \texttt{Englisch}\} \times \{\texttt{Spam}, \texttt{keinSpam}\},$$

$$p\big((\texttt{Deutsch}, \texttt{Spam})\big) = 0.2,$$
$$p\big((\texttt{Deutsch}, \texttt{keinSpam})\big) = 0.1,$$
$$p\big((\texttt{Englisch}, \texttt{Spam})\big) = 0.6,$$
$$p\big((\texttt{Englisch}, \texttt{keinSpam})\big) = 0.1$$

4. (Anzahl Würfe, bevor beim wiederholten fairen Münzwurf zum ersten Mal Kopf kommt) $\Omega = \mathbb{N}_0$, $p(n) = (1/2)^n \cdot (1/2) = 2^{-n-1}$ für $n \in \Omega$

Bemerkung 1.4

Wahrscheinlichkeitsräume $(\Omega, \mathcal{F}, P)$ gemäß Def. 1.1 werden heute, nicht zuletzt wegen des einflussreichen Werks von A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1933, oft als „Basisobjekt“ der mathematischen Modellierung von Zufallsvorgängen verwendet. Insoweit ist die explizite Wahl eines geeigneten Wahrscheinlichkeitsraums (oft) ein Teil der Arbeit, wenn ein umgangssprachlich formuliertes Problem in Mathematik übersetzt werden soll.
Man sollte sich bewusst sein, dass es dabei i.A. viele mögliche Wahlen gibt und dass die Formulierung mit Zufallsvariablen oftmals einen für die Intuition und für das Argumentieren sehr angenehmen Zugang ergibt (s.u.).

Siehe auch Abschnitt 1.2 („Kleingedrucktes“) der Vorlesungsnotizen für Hintergrund und weitere Diskussion.

Lemma 1.5

Für $A, B, A_1, A_2, \dots \in \mathcal{F}$ gilt

$$P(\varnothing) = 0 \tag{3}$$

$$P(A \cup B) + P(A \cap B) = P(A) + P(B) \quad \textit{(endliche Additivität)} \tag{4}$$

$$\textit{insbesondere } P(A) + P(A^c) = 1$$

$$A \subset B \Rightarrow \; P(A) \le P(B) \qquad \textit{(Monotonie)} \tag{5}$$

zudem gilt auch

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) \le \sum_{n=1}^{\infty} P(A_n) \qquad (\sigma\textit{-Subadditivität}) \tag{6}$$

falls $A_n \nearrow_{n\to\infty} A$ *(d.h.* $A_1 \subset A_2 \subset \dots$ *mit* $A = \bigcup_{n=1}^{\infty} A_n$*)*
oder $A_n \searrow_{n\to\infty} A$ *(d.h.* $A_1 \supset A_2 \supset \dots$ *mit* $A = \bigcap_{n=1}^{\infty} A_n$*),*
so gilt $P(A) = \lim_{n\to\infty} P(A_n)$ $\qquad$ *(σ-Stetigkeit)* $\quad$ (7)

`[Argument ggf. an der Tafel]`

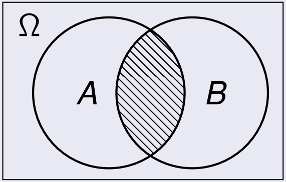
Bemerkung 1.6 (Einschluss-Ausschluss-Formel)

Formel (4) aus Lemma 1.5 kann man auch folgendermaßen schreiben

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

Man kann diese Formel anhand eines Venn-Diagramms veranschaulichen:

Der schraffierte Bereich $A \cap B$ wird in $P(A) + P(B)$ doppelt gezählt, in $P(A \cap B)$ aber nur einmal.



Bemerkung 1.6 (allgemeine Form der Einschluss-Ausschluss-Formel, „Siebformel“)

Allgemein gilt für $A_1, A_2, \ldots, A_n \in \mathcal{F}$

$$
\begin{aligned}
P\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n P(A_i) - \sum_{\substack{i,j=1\\ i\neq j}}^n P(A_i \cap A_j) \\
+ \sum_{\substack{i,j,k=1\\ i\neq j, i\neq k, j\neq k}}^n P(A_i \cap A_j \cap A_k) \\
\pm \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n)
\end{aligned}
$$

d.h. knapp ausgedrückt

$$
P\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{\varnothing \neq J \subset \{1,\ldots,n\}} (-1)^{(\#J)-1} P\Big(\bigcap_{i\in J} A_i\Big) \tag{8}
$$

[z.B. induktiv, siehe auch Notizen, Abschnitt 1.2.3]

Beispiel

Betrachten wir in Bsp. 1.3, 1 (d) mit $n = 3$ die Ereignisse $A_i = \{(x_1, x_2, x_3) \in \Omega : x_i = i\}$ für $i = 1, 2, 3$. Wenn wir eine uniform verteilte Permutation von $1, 2, 3$ betrachten, so ist $A_i = \{i \text{ ist ein Fixpunkt}\}$. Dann ist

$$
\begin{aligned}
&P(A_1 \cup A_2 \cup A_3) \\
&= P(A_1) + P(A_2) + P(A_3) \\
&\qquad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\
&\qquad + P(A_1 \cap A_2 \cap A_3) \\
&= \frac{1}{3} + \frac{1}{3} + \frac{1}{3} - \frac{1}{6} - \frac{1}{6} - \frac{1}{6} + \frac{1}{6} = \frac{2}{3}
\end{aligned}
$$

die Wahrscheinlichkeit, dass es mindestens einen Fixpunkt gibt.

Bemerkung 1.7

Wir lassen hier die „philosophische“ Frage offen, welche Interpretation die Wahrscheinlichkeit $P(A)$ eines Ereignisses A haben soll. Es bieten sich etwa an:

- „naive“ Interpretation der Wahrscheinlichkeit: Die „Natur“ enthält inhärente Unbestimmtheit („ist selbst nicht sicher, was sie tut“), und $P(A)$ beschreibt den Grad der Sicherheit, mit dem sie sich für das Ereignis A entscheidet.
- „frequentistische“ Interpretation der Wahrscheinlichkeit: Wenn man das zufällige Experiment unter exakt denselben Bedingungen sehr oft wiederholte, wäre der relative Anteil der Ausgänge, in denen A eingetreten ist, etwa $P(A)$.
- „subjektive“ Interpretation der Wahrscheinlichkeit: $P(A)$ misst, wie sicher ich mir persönlich bin, dass A eintreten wird. (Beispielsweise: Wieviel Wetteinsatz wäre ich bereit zu bezahlen, wenn mir 1 € ausgezahlt würde, sofern A eintritt?)

Bemerkung 1.7 (Fortsetzung)

Kersting & Wakolbinger [KW, S. vi] schreiben dazu: „Es kann nicht darum gehen, eine spezielle Intuition gegenüber den anderen durchzusetzen. Dass sich dies innerhalb der Mathematik auch gar nicht als nötig erweist, ist eine der Stärken mathematischer Wissenschaft."

Siehe z.B. auch die Diskussion in Georgii [Ge, S. 14] am Ende von Abschn. 1.1.3.

Zufallsvariablen

Oft stellt man sich zumindest intuitiv Zufallsvorgänge so vor, dass ein „zufälliger Wert“ beobachtet oder gemessen wurde, mit dem Verständnis, dass sich bei Wiederholung des Experiments möglicherweise ein anderer Wert aus einer gewissen Menge möglicher Werte ergäbe.

So kann man die Beispiele 1.3 auch folgendermaßen aussprechen:

1. (a) Sei W das Ergebnis eines Wurfs eines fairen 6er-Würfels (Wertebereich $\{1, 2, 3, 4, 5, 6\}$).
   (b) Werfe Würfel dreimal, seien W_1, W_2, W_3 Ergebnisse des 1., 2., 3. Wurfs des Würfels ($W = (W_1, W_2, W_3)$ hat Wertebereich $\{1, 2, 3, 4, 5, 6\}^3$).
   (c) $Z = (X, Y)$ ein zufällig gewähltes Pixel aus $[0, 1]$ (aus Auftakt-Beispiel aus Kapitel 0 mit Diskretisierung auf 32 Bit Genauigkeit)

So kann man die Beispiele 1.3 auch folgendermaßen aussprechen:

1. (d) n verschiedene Eingaben in zufälliger Reihenfolge: Sei $X = (X_1, X_2, \ldots, X_n)$ eine zufällige Permutation von $1, 2, \ldots, n$ (Wertebereich $\{(x_1, x_2, \ldots, x_n) : x_1, \ldots, x_n \in \{1, 2, \ldots, n\}$ paarweise verschieden$\}$)
2. Sei M das Ergebnis eines (möglicherweise verfälschten) Münzwurfs (Wertebereich $\{\mathrm{Kopf}, \mathrm{Zahl}\}$)
3. Wähle uniform eine Email aus meiner Inbox (die deutsche und englische Emails enthält, die jeweils Spam sein können oder nicht), sei X die Sprache und Y der Spam-Status der betrachteten Email ((X, Y) hat Wertebereich $\{\mathtt{Deutsch}, \mathtt{Englisch}\} \times \{\mathtt{Spam}, \mathtt{keinSpam}\}$)
4. Wir werfen eine faire Münze, bis zum ersten Mal Kopf kommt. G zählt, wie oft bis dahin Zahl gefallen ist.

Definition 1.8

$(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, S eine Menge, eine Abbildung* $X : \Omega \to S$ nennen wir eine *Zufallsvariable*, oft abgekürzt ZV (auch: *Zufallsgröße*) mit Wertebereich S.

ZVn und Ereignisse: Für $A \subset S$ schreibt man*

$$\{X \in A\} := \{\omega \in \Omega : X(\omega) \in A\} = X^{-1}(A)$$

für das Ereignis „X nimmt einen Wert in A an", für $B = \{x\}$ mit $x \in S$ abkürzend auch oft $\{X = x\} := \{X \in \{x\}\}$; im Fall $S = \mathbb{R}$ oft auch $\{X \leq x\} := \{X \in (-\infty, x]\}$, etc.

Beachte hier die übliche Notationskonvention: ZV werden meist mit Großbuchstaben benannt, mögliche Werte („Realisierungen") mit Kleinbuchstaben.

---

* Im Fall, dass S überabzählbar ist, muss man strenggenommen *Messbarkeit* fordern, siehe Notizen, Abschnitt 1.2.1.

Beispiel 1.9 (Indikatorvariable)

$A \in \mathcal{F}$, $1_A : \Omega \to \{0, 1\}$,

$$1_A(\omega) = \begin{cases} 1, & \text{falls } \omega \in A, \\ 0, & \text{falls } \omega \notin A \end{cases}$$

1_A heißt die Indikatorvariable des Ereignisses A.

Verteilung

Definition 1.10

Für eine Zufallsvariable X mit Wertebereich S heißt

$$\rho_X : A \mapsto P(X \in A), \quad A \subset S$$

die *Verteilung* von X.

Eine ZV X heißt *diskret*, wenn ihr Wertebereich S (endlich oder) abzählbar ist oder zumindest eine (endliche oder) abzählbare Teilmenge S enthält mit $P(X \in S) = 1$. Die Zahlen

$$\rho_X(\{a\}) := P(X = a), \quad a \in S$$

heißen dann die *Verteilungsgewichte* von X (oft auch nur: die *Gewichte*).

Wir kürzen oft ab (mit einem kleinen „Notationsmissbrauch“) $\rho_X(a) := \rho_X(\{a\})$.

Verteilung (Fall mit Gewichten)

$$\rho_X(A) = P(X \in A),\ A \subset S, \quad \rho_X(a) = P(X = a),\ a \in S$$

Offenbar gilt (mit Eigenschaft (A) aus Forderung (2)):
$\rho_X(a) \geq 0$,

$$\sum_{a \in A} \rho_X(a) = P(X \in A) \text{ für } A \subset S, \qquad \text{insbes. } \sum_{a \in S} \rho_X(a) = 1$$

und ρ_X ist ein *Wahrscheinlichkeitsmaß* (auch *Wahrscheinlichkeitsverteilung*) auf S (d.h. $\rho_X : \{\text{Teilmengen von } S\} \to [0,1]$ und die (analogen) Eigenschaften aus Forderung (2) und aus Lemma 1.5 gelten, mit Lesung $\Omega = S$).

Man schreibt für die Verteilung einer ZV X oft auch $\mathscr{L}(X)$ (das $\mathscr{L}$ erinnert an English "law" bzw. Französisch ≪loi≫, d.h. „Gesetz"), man schreibt auch $X \sim \rho$, wenn $\mathscr{L}(X) = \rho$ gilt (für eine gewisse Wahrscheinlichkeitsverteilung).

Beobachtung 1.11 (Operationen mit Zufallsvariablen)

Zufallsvariablen sind nicht zuletzt deshalb nützlich für die Modellierung zufälliger Vorgänge, weil man mit ihnen gewissermaßen operieren und „rechnen" kann wie mit den Variablen in einer Programmiersprache:

Seien X und Y mit Wertebereichen S_X bzw. S_Y ZVn auf demselben Wahrscheinlichkeitsraum, $f : S_X \to S$ eine Abbildung, so können wir die S-wertige ZV

$$f(X) := f \circ X$$

bilden und die $S_X \times S_Y$-wertige ZV (X, Y).

Speziell können wir im Fall $S_X, S_Y \subset \mathbb{R}$ die ZVn X^2, $X + Y$, $X - Y$, XY, etc. bilden.

Beobachtung 1.11 (Operationen mit Zufallsvariablen, 2)

Mit $f : S_X \to S$ wie oben ergibt sich (im diskreten Fall) für die Gewichte von $f(X)$ dann

$$\begin{aligned}\rho_{f(X)}(s) &= P\big(f(X) = s\big) = P\big(X \in \{x \in S_X : f(x) = s\}\big) \\ &= \sum_{x\,:\,f(x)=s} P(X = x) = \sum_{x\,:\,f(x)=s} \rho_X(x)\end{aligned}$$

für $s \in S$.

Beispiel (Verteilung der Augensumme zweier fairer Würfel)

Sei (W_1, W_2) das Ergebnis des 1. und 2. Wurf eines fairen Würfels,
$f : \{1, 2, \dots, 6\}^2 \to \{2, 3, \dots, 12\}, (w_1, w_2) \mapsto w_1 + w_2$, so ist

$$\begin{aligned}
P(W_1 + W_2 = s) &= P(f(W_1, W_2) = s) \\
&= \sum_{(w_1, w_2) : w_1 + w_2 = s} P(W_1 = w_1, W_2 = w2) \\
&= \frac{\#\{(w_1, w_2) \in \{1, 2, \dots, 6\}^2 : w_1 + w_2 = s\}}{36} \\
&= \frac{6 - |7 - s|}{36}
\end{aligned}$$

für $s \in \{2, 3, \dots, 12\}$.

Hinweis

Bemerkung 1.12 (Kanonisches Modell für eine ZV)

Man kann eine Zufallsvariable X mit Wertebereich S und Verteilung ρ stets in „kanonischer Weise" mit der Wahl $\Omega = S$ und $P = \rho$ und geeignetem $\mathcal{F} \subset 2^S$ (im diskreten Fall kann man $\mathcal{F} = 2^S$ wählen) als $X = \mathrm{Id}_S$ auf dem W'raum $(S, \mathcal{F}, \rho)$ formulieren.

Man kann daher genausogut die mathematische Modellierung eines Zufallsphänomens mit der Formulierung geeigneter Zufallsvariablen samt Verteilung beginnen, siehe auch Diskussion und Referenzen in Abschnitt 1.2.2 der Notizen.

Urnenmodelle

Beispiel 1.13 (Ziehen mit und ohne Zurücklegen: Urnenmodelle)

Eine Urne enthalte n (mit $1, 2, \ldots, n$ nummerierte) Kugeln, wir ziehen zufällig k ($\leq n$) heraus. Sei

$$X_i \text{ die Nummer der Kugel im } i\text{-ten Zug}, \quad i = 1, \ldots, k$$

und $X = (X_1, X_2, \ldots, X_k)$.

n Kugeln in Urne,

X_i die Nummer der Kugel im i-ten Zug, $\quad i = 1, \ldots, k$

Wir betrachten 4 mögliche Situationen:

1. Ziehen mit Zurücklegen, mit Beachtung der Reihenfolge:

X hat Werte in
$W_1 = \big\{(x_1, \ldots, x_k) : x_1, \ldots, x_k \in \{1, 2, \ldots, n\}\big\} = \{1, 2, \ldots, n\}^k$,

$$P\big(X = (x_1, \ldots, x_k)\big) = \frac{1}{n^k}$$

(d.h. X ist uniform auf $\{1, \ldots, n\}^k$ verteilt).

n Kugeln in Urne,

X_i die Nummer der Kugel im i-ten Zug, $\quad i = 1, \dots, k$

2. Ziehen ohne Zurücklegen, mit Beachtung der Reihenfolge:

X hat Werte in

$W_2 = \{(x_1, \dots, x_k) : x_1, \dots, x_k \in \{1, 2, \dots, n\}, x_i \neq x_j \text{ für } i \neq j\}$

$$P\big(X = (x_1, \dots, x_k)\big) = \frac{1}{n \cdot (n-1) \cdots (n-k+1)}$$
$$= \frac{(n-k)!}{n!} \quad \left(= \frac{1}{|W_2|} \right)$$

3. Ziehen ohne Zurücklegen, ohne Beachtung der (Zug-)Reihenfolge: Wir beobachten nicht X (das hier wie in 2. verteilt wäre), sondern mit

$$\varphi\big((x_1,\dots,x_k)\big) = \{x_1,\dots,x_k\}$$ (verwandle in Menge, d.h. vergiss Reihenfolge)

(nur) $Y = \varphi(X)$ mit Werten in $W_3 = \big\{A \subset \{1,2,\dots,n\} : \#A = k\big\}$.

Es ist

$$P(Y = A) = \frac{1}{\binom{n}{k}} = \frac{k!(n-k)!}{n!} \quad \left(= \frac{1}{|W_3|},\ \text{es gibt } \tbinom{n}{k} \text{ versch. } k\text{-elementige Teilmengen} \right)$$

denn

$$P(Y = A) = P\big(\varphi(X) = A\big) = P\big(X \in \varphi^{-1}(A)\big)$$
$$= \sum_{x \in W_2 : \varphi(x) = A} P(X = x) = \sum_{x \in W_2 : \varphi(x) = A} \frac{(n-k)!}{n!} = k! \frac{(n-k)!}{n!}$$

(es gibt $k!$ viele verschiedene Elemente x von W_2 mit $\varphi(x) = A$, nämlich alle verschiedenen Anordnungen der k Elemente von A).

4. Ziehen mit Zurücklegen, ohne Beachtung der (Zug-)Reihenfolge:

Sei $X = (X_1, \dots, X_k)$ wie in 1., wir beobachten aber (nur) $Y = (Y_1, \dots, Y_n)$, wobei

$$Y_j = \#\{1 \le i \le k : X_i = j\} \quad \text{für } j = 1, \dots, n.$$

(Y_j gibt an, wie oft Kugel j gezogen wurde). Y hat Werte in

$$W_4 = \Big\{(\ell_1, \ell_2, \dots, \ell_n) \in \mathbb{N}_0^n : \ell_1 + \ell_2 + \dots + \ell_n = k\Big\}$$

Man nennt Y auch einen (zufälligen) „Besetzungsvektor".

Für $(\ell_1, \ell_2, \dots, \ell_n) \in W_4$ gibt es

$$\binom{k}{\ell_1, \ell_2, \dots, \ell_n} := \frac{k!}{\ell_1! \cdot \ell_2! \cdot \dots \cdot \ell_n!} \qquad \text{„Multinomialkoffizient"}$$

verschiedene $x = (x_1, \dots, x_k) \in W_1$ mit

$$\Big|\{1 \le i \le k : x_i = j\}\Big| = \ell_j \quad \text{für } j = 1, \dots, n.$$

`[Ggf. Argument an der Tafel]`

4. Ziehen mit Zurücklegen, ohne Beachtung der (Zug-)Reihenfolge:

Sei $X = (X_1, \dots, X_k)$ wie in 1., wir beobachten aber (nur) $Y = (Y_1, \dots, Y_n)$, wobei

$$Y_j = \#\{1 \le i \le k : X_i = j\} \quad \text{für } j = 1, \dots, n$$

mit Werten in

$$W_4 = \Big\{(\ell_1, \ell_2, \dots, \ell_n) \in \mathbb{N}_0^n : \ell_1 + \ell_2 + \dots + \ell_n = k\Big\}$$

Somit ist

$$P\big(Y = (\ell_1, \dots, \ell_n)\big) = \binom{k}{\ell_1, \ell_2, \dots, \ell_n}\Big(\frac{1}{n}\Big)^k, \quad (\ell_1, \dots, \ell_n) \in W_4$$

d.h. eine Instanz der Multinomialverteilung

(wir kommen in Bsp. 1.18 darauf zurück).

Bemerkung 1.14

Es gilt

$$|W_4| = \binom{n+k-1}{k} \quad \left(= \binom{n+k-1}{n-1} \right)$$

Ein “Zähltrick”: Lege k Kugeln und $n-1$ “Trennstäbe” – also insgesamt $n+k-1$ Objekte – in eine Reihe:

$$\underbrace{\bigcirc\cdots\bigcirc}_{\ell_1 \text{ Kugeln}} \Big| \underbrace{\bigcirc\bigcirc\cdots\bigcirc}_{\ell_2 \text{ Kugeln}} \Big| \underbrace{}_{\ell_3 = 0} \Big| \cdots \Big| \underbrace{\bigcirc\bigcirc\cdots\bigcirc}_{\ell_{n-1} \text{ Kugeln}} \Big| \underbrace{\bigcirc\cdots\bigcirc}_{\ell_n \text{ Kugeln}}$$

Insbesondere ist die Verteilung auf W_4 aus Beispiel 1.13, 4. nicht die uniforme.

Die uniforme Verteilung auf dem W_4 aus Beispiel 1.13, 4. heißt auch die „Bose-Einstein-Verteilung“, die in Beispiel 1.13, 4. betrachtete Verteilung heißt die „Maxwell-Boltzmann-Verteilung“.

Beispiel. Eine Hörsaalreihe habe n Plätze, darauf nehmen m ($\leq n/2$) Männer und $n - m$ Frauen rein zufällig Platz.

Die Wahrscheinlichkeit, dass keine zwei Männer nebeneinander sitzen

$$= \frac{\binom{n-m+1}{m}}{\binom{n}{m}}$$

Zum Zähler: Setze m Männer auf $n - m + 1$ Plätze, dann setze jeweils eine Frau rechts des i-ten Manns für $i = 1, 2, \dots, m - 1$

z.B. für $n = 17$, $m = 5$ ist $\binom{17}{5} = 6188$, $\binom{13}{5} = 1287$, $\binom{13}{5}/\binom{17}{5} \approx 0{,}208$

Beispiel 1.15 (Hypergeometrische Verteilung)

Eine Urne enthalte n Kugeln, davon s schwarze und w weiße ($s + w = n$), ziehe k-mal ohne Zurücklegen,

$$\mathrm{Hyp}_{s,w,k}\big(\{\ell\}\big) = \frac{\binom{s}{\ell}\binom{w}{k-\ell}}{\binom{s+w}{k}}, \quad \ell = 0, 1, \ldots, k$$

ist die W'keit, genau ℓ schwarze Kugeln zu ziehen.

Z.B. ist die W'keit, dass der Geber beim Skat genau 3 Asse bekommt $= \dfrac{\binom{4}{3}\binom{28}{7}}{\binom{32}{10}} \approx 0{,}073$.

Beispiel 1.16 (p-Münzwurf und Binomialverteilung)

1. $S = \{0,1\}$, $\mathrm{Ber}_p(\{1\}) = p = 1 - \mathrm{Ber}_p(\{0\})$ mit einem $p \in [0,1]$ („Bernoulli-Verteilung“[a])

2. n-facher p-Münzwurf (mit $p \in [0,1]$): $S = \{0,1\}^n$,

$$\mathrm{Ber}_p^{\otimes n}\big(\{(x_1,\dots,x_n)\}\big) = p^{|\{i \le n : x_i = 1\}|}(1-p)^{|\{i \le n : x_i = 0\}|}$$

Für $X = (X_1,\dots,X_n) \sim \mathrm{Ber}_p^{\otimes n}$ sagt man: $X_1,\dots,X_n$ unabhängig und $X_i \sim \mathrm{Ber}_p$ für $i = 1,2,\dots,n$

3. Binomialverteilung (zum Parameter n und p, $n \in \mathbb{N}$, $p \in [0,1]$):

$$\mathrm{Bin}_{n,p}\big(\{k\}\big) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in S := \{0,1,\dots,n\}$$

(dies ist die W'keit, beim n-fachen Münzwurf genau k Erfolge zu beobachten)

Beispiel 1.16, 3. mit Zufallsvariablen ausgesprochen:

$$X = (X_1, \dots, X_n) \sim \mathrm{Ber}_p^{\otimes n}, \qquad Y := X_1 + X_2 + \dots + X_n,$$

so ist $Y \sim \mathrm{Bin}_{n,p}$, denn für $k \in \{0, 1, \dots, n\}$ und für jede Wahl von $(x_1, \dots, x_n) \in \{0, 1\}^n$ mit $x_1 + \dots + x_n = k$ ist $P(X_1 = x_1, \dots, X_n = x_n) = p^k (1-p)^{n-k}$, somit mit $B_k = \big\{(x_1, \dots, x_n) \in \{0, 1, \}^n : x_1 + \dots + x_n = k\big\}$

$$\begin{aligned} P(Y = k) &= P\Big((X_1, \dots, X_n) \in B_k\Big) \\ &= \sum_{(x_1, \dots, x_n) \in B_k} P\Big((X_1, \dots, X_n) = (x_1, \dots, x_n)\Big) \\ &= \sum_{(x_1, \dots, x_n) \in B_k} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

da $\# B_k = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = \binom{n}{k}$.

Insoweit ist dies eine Beispiel-Instanz zu Beobachtung 1.11.

Beispiel 1.17 (Geometrische Verteilung)

$p \in (0, 1)$, $S = \mathbb{N}_0$,

$$\mathrm{Geom}_p\big(\{j\}\big) = p(1-p)^j, \quad j \in \mathbb{N}_0$$

ist die W'keit, bei wiederholtem p-Münzwurf genau k Misserfolge vor dem ersten Erfolg zu beobachten.

Beachte: Manche Autoren betrachten die geometrische Verteilung auf $\mathbb{N}$ (statt auf $\mathbb{N}_0$), dann ist das Gewicht $p(1-p)^{k-1}$ und die Interpretation „k Würfe (einschließlich) bis zum ersten Erfolg."

Beispiel 1.18 (Multinomialverteilung)

$s \in \{2, 3, \dots\}$, $p_1, \dots, p_s \in [0, 1]$, $p_1 + \dots + p_s = 1$, $n \in \mathbb{N}$,
$S = \{(k_1, \dots, k_s) \in \mathbb{N}_0^s : k_1 + \dots + k_s = n\}$,

$$\mathrm{Mult}_{n;p_1,\dots,p_s}\big(\{(k_1, \dots, k_s)\}\big) = \binom{n}{k_1, k_2, \dots, k_s} p_1^{k_1} p_2^{k_2} \cdots p_s^{k_s}$$

Interpetation: n Züge mit Zurücklegen ohne Beachtung der Reihenfolge aus einer Urne mit s Kugeln (s verschiedene „Farben“), Farbe i wird mit W'keit p_i gezogen), obiges ist die W'keit, genau k_i-mal Farbe i zu ziehen für $i = 1, 2, \dots, s$.

Beispiel 1.19 (Poissonverteilung*)

$\lambda \in (0, \infty)$,

$$\mathrm{Poi}_\lambda\big(\{k\}\big) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0$$

---

*nach Siméon Denis Poisson, 1781–1840

Proposition 1.20 (Poissonapproximation der Binomialverteilung)

Seien $p_n \in [0,1]$ mit $p_n \to 0$ und $np_n \to \lambda \in (0,\infty)$ für $n \to \infty$, so gilt für jedes $k \in \mathbb{N}_0$

$$\mathrm{Bin}_{n,p_n}(\{k\}) \underset{n\to\infty}{\longrightarrow} \mathrm{Poi}_\lambda(\{k\}).$$

Beweis. Es ist

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} = \underbrace{\frac{n(n-1)\cdots(n-k+1)}{k!\,n^k}}_{\to 1/k!} \big(\underbrace{np_n}_{\to\lambda}\big)^k \underbrace{\left(1-\frac{np_n}{n}\right)^n}_{\to e^{-\lambda}} (1-p_n)^{-k}$$

$$\longrightarrow e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{für } n\to\infty.$$

Prop. 1.20:

$$np_n \to \lambda \in (0, \infty), \quad \text{so gilt} \quad \mathrm{Bin}_{n,p_n}(k) \underset{n\to\infty}{\longrightarrow} \mathrm{Poi}_\lambda(k)$$

Prop. 1.20 motiviert, warum die Poissonverteilung oft in Anwendungssituationen vorkommt, in denen man viele unabhängige Ereignisse betrachtet, von denen jedes nur mit einer sehr kleinen W'keit eintritt.

Man denke etwa an Schädensfälle bei Versicherungen, Zerfallsereignisse in einer Probe radioaktiven Materials oder an genetische Mutationen.

Beispiel 1.21

L. von Bortkewitsch berichtete in seinem Buch *Das Gesetz der kleinen Zahlen*, Teubner, 1898 verschiedene Datensätze, die gut zur Poissonverteilung passen.

Beispiel 1.21

Speziell in § 12, 4. („Die durch Schlag eines Pferdes im preußischen Heere getöteten") werden für 20 Jahre (1875–1894) und 10 Armeekops der preußischen Kavallerie, also insgesamt $20 \cdot 10 = 200$ „Korpsjahre" berichtet, in wievielen davon sich x Todesfälle durch Schlag eines Pferds ereigneten (Tabelle b) auf S. 25):

Ergebnis x	Anz. „Korpsjahre"
0	109
1	65
2	22
3	3
4	1
≥ 5	0

Beispiel 1.21, Forts.

Angenommen, die Anzahl durch Schlag eines Pferdes während eines Jahres in einem Korps getöteter Soldaten wäre Poi_λ-verteilt mit $\lambda = 0{,}61$, so würden wir das Resultat x je $200 \times \mathrm{Poi}_{0,61}(x)$-mal erwarten:

Ergebnis x	Anz. „Korpsjahre"	$200 \times \mathrm{Poi}_{0,61}(x)$
0	109	108,67
1	65	66,29
2	22	20,22
3	3	4,11
4	1	0,63
≥ 5	0	0,08

Von Bortkewitsch, a.a.O., S. 25 schreibt: „Die Kongruenz der Theorie mit der Erfahrung lässt [...], wie man sieht, nichts zu wünschen übrig."

Beispiel 1.21, Forts. 2

Übrigens: Wie ist von Bortkewitsch auf $\lambda = 0{,}61$ gekommen?

Die beobachtete „mittlere Anzahl Todesfälle pro Korpsjahr“ in den Daten ist

$$\widehat{\lambda} = \frac{109}{200} \cdot 0 + \frac{65}{200} \cdot 1 + \frac{22}{200} \cdot 2 + \frac{3}{200} \cdot 3 + \frac{1}{200} \cdot 4 + 0 = 0{,}61$$

und es ist auch

$$\sum_{x=0}^{\infty} x \mathrm{Poi}_\lambda(x) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} = \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda} = \lambda$$

(„der Erwartungswert von Poi_λ ist λ“) und somit ist obiges der naheliegende „Momentenschätzer“ – wir werden darauf zurückkommen.