Beispiel 1.22 (Negative Binomialverteilung, auch Pascal-Verteilung genannt)**.** Für $r \in (0,\infty)$, $p \in (0,1)$ ist

$$\mathrm{NegBin}_{r,p}\big(\{k\}\big) = \binom{-r}{k}(-1)^k p^r (1-p)^k, \qquad k \in \mathbb{N}_0$$

wobei $\binom{-r}{k} := \frac{(-r)(-r-1)\cdots(-r-k+1)}{k!}$ $\Big(= (-1)^k \binom{r+k-1}{k}$ für $r \in \mathbb{N}\Big)$.

$\mathrm{NegBin}_{r,p}\big(\{k\}\big)$ ist die W'keit, insgesamt k Misserfolge vor dem r-ten Erfolg in einer p-Münzwurffolge zu sehen ($\mathrm{NegBin}_{1,p} = \mathrm{Geom}_p$).

(Beachte

$$\sum_{k=0}^{\infty} \binom{-r}{k}(-1)^k p^r (1-p)^k = p^r \big(1 + p - 1\big)^{-r} = 1$$

und allg. für $r > 0$ $\sum_{k=0}^{\infty} \binom{-r}{k} x^k = (1+x)^{-r}$ (für $|x| < 1$) mittels Taylor-Entwicklung in $x = 0$ bzw. allgemeinem Binomischem Lehrsatz.)

Beispiel 1.23 (Multinomialverteilung)**.** $s \in \{2,3,\dots\}$, $p_1,\dots,p_s \in [0,1]$, $p_1 + \cdots + p_s = 1$, $n \in \mathbb{N}$,
$\Omega = \big\{(k_1,\dots,k_s) \in \mathbb{N}_0^s : k_1 + \cdots + k_s = n\big\}$,

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

$= 1^n$

$= 1$

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 $i = 1,2,\dots,s$.

$\Big(\sum_{(k_1,\dots,k_s)\in\Omega} \mathrm{Mult}_{n;p_1,\dots,p_s}(\{(k_1,\dots,k_s)\}) = (p_1+p_2+\dots+p_s)^n$

Beispiel 1.24 (Poissonverteilung[8]). $\lambda \in (0, \infty)$,

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

(beachte: $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$)

Proposition 1.25 (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* $k \in \mathbb{N}_0$

$$\mathrm{Bin}_{n,p_n}(\{k\}) \xrightarrow[n\to\infty]{} \mathrm{Poi}_\lambda(\{k\}).$$

(z. B.: $p_n = \frac{\lambda}{n}$)

denn: $\mathrm{Bin}_{n,p_n}(\{k\}) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \underbrace{\frac{n(n-1)\cdots(n-k+1)}{k!\, n^k}}_{\xrightarrow[n\to\infty]{} \frac{1}{k!}} \underbrace{(np_n)^k}_{\xrightarrow[n\to\infty]{} \lambda^k} \underbrace{\left(1-\frac{np_n}{n}\right)^n}_{} \cdot \underbrace{(1-p_n)^{-k}}_{\xrightarrow[n\to\infty]{} 1}$

$$= \left(1-\frac{np_n}{n}\right)^n \sim \left(1-\frac{\lambda}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-\lambda}$$

Prop. 1.25 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.

[8] nach Siméon Denis Poisson, 1781–1840

Beispiel 1.26. L. von Bortkewitsch[9] berichtete in seinem Buch *Das Gesetz der kleinen Zahlen*, Teubner, 1898 verschiedene Datensätze, die gut zur Poissonverteilung passen.

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

[9] Ladislaus von Bortkewitsch, 1868–1931

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.“

Warum $\lambda = 0{,}61$?

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

ist die „mittlere Anz. Todesfälle pro Korpsjahr“ (in den Daten),

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

1.4 Zufallsvariablen

Zufallsvariablen (oft abgekürzt ZV; manchmal auch Zufallsgrößen gennant) sind (in einem gewissen Sinn sogar: die) „Fundamentalobjekte“ der Stochastik[10], sie sind zudem oft sehr angenehm zum Rechnen/Notieren und zum intuitiven Argumentieren über zufällige Vorgänge.

So kann man die Beispiel-Instanzen aus Beispiel 1.4 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) n verschiedene Objekte 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 $\{\texttt{Deutsch}, \texttt{Englisch}\} \times \{\texttt{Spam}, \texttt{keinSpam}\}$)
4. Wir werfen eine faire Münze, bis zum ersten Mal Kopf kommt. G zählt, wie oft bis dahin Zahl gefallen ist (Wertebereich $\mathbb{N}_0$).

[10] Man kann ein Zufallsexperiment auch stets auffassen als: „Der Zufall wählt aus einer Menge S möglicher Realisierungen der Zufallsvariable X eine aus“ und sich in diesem Sinn den „Umweg“ über die Diskussion von Ereignissen und Wahrscheinlichkeitsräume (zunächst) ersparen, tatsächlich stellt das Buch von Kersting & Wakolbinger [KW] Zufallsvariablen auch gleich an den Anfang der Diskussion; die beiden Zugänge sind logisch äquivalent, unterscheiden sich aber etwas in der „Betonung“. Siehe auch Bemerkung 1.36 unten.

Beispiel 1.27. Wir werfen 2 Würfel (naheliegende Modellierung: $\Omega = \{1, 2, \dots, 6\}^2$ [und $\mathcal{F} = 2^\Omega$, P=uniforme Verteilung auf Ω]) und beobachten die Augensumme, nennen wir sie X.
Wir könnten übergehen zu $\Omega' = \{2, 3, \dots, 12\}$ [und entsprechendem $P'(\{x\}) = \frac{6-|x-7|}{36}$] oder wir betrachten

$$\begin{array}{rccc} X : & \Omega & \to & \Omega' \\ & \uplus & & \\ \omega = & (\omega_1, \omega_2) & \mapsto & \omega_1 + \omega_2 \end{array}$$

und Ereignisse

$$\{X = x\} := \{(\omega_1, \omega_2) \in \Omega : \omega_1 + \omega_2 = x\} \quad \text{für } x = 2, 3, \dots, 12.$$

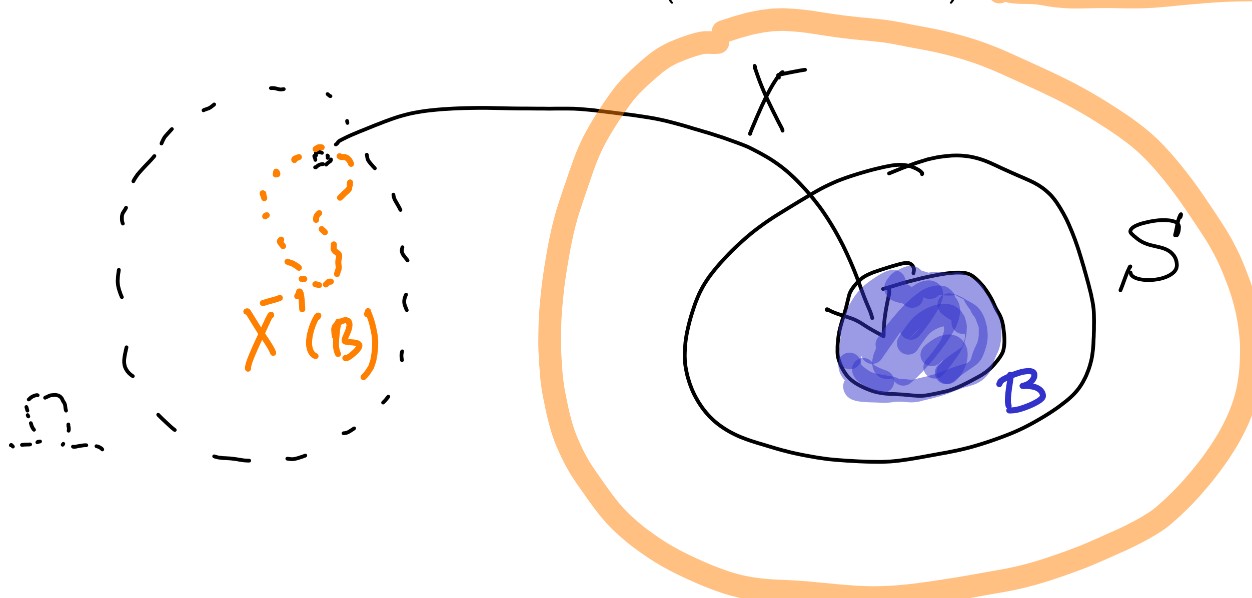
Nehmen wir an, ein anderer Beobachter bemerkt (Augenzahl 1. Wurf) − (Augenzahl 2. Wurf), dieser müsste zu $\Omega'' = \{-5, -4, \dots, 5\}$ und P'' übergehen, oder aber

$Y : \Omega \to \Omega''$ mit $Y((\omega_1, \omega_2)) = \omega_1 - \omega_2$ betrachten.

Definition 1.28. $(\Omega, \mathcal{F}, P)$ ein W'raum, $(S, \mathcal{S})$ ein messbarer Raum (der „Wertebereich").
Eine Abbildung $X : \Omega \to S$ heißt eine *Zufallsvariable* (mit Wertebereich S), wenn gilt

$$\forall\, B \in \mathcal{S} : X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

(im Sinne der Maßtheorie ist X eine messbare (strikt: eine $\mathcal{F}$-$\mathcal{S}$)-messbare Abbildung von Ω nach S.)



Wir schreiben

$$\{X \in B\} := X^{-1}(B)$$

für das Ereignis „X nimmt einen Wert in B an" und abkürzend oft auch $\{X = x\} := \{X \in \{x\}\}$; im Fall $S = \mathbb{R}$ oft auch $\{X \le x\} := \{X \in (-\infty, x]\}$, etc.
Im Fall reellwertiger Zufallsvariablen lassen wir oft auch die Werte $+\infty$ oder $-\infty$ zu (Übergang von $S = \mathbb{R}$ zu $S = \overline{\mathbb{R}}$).

Übliche Notationskonvention: ZV werden meist mit Großbuchstaben benannt, mögliche Werte („Realisierungen") mit Kleinbuchstaben.

Bericht 1.29. 1. Wenn Ω und S abzählbar sind (und wir in kanonischer Weise die jeweiligen Potenzmengen als σ-Algebren verwenden), so ist jede Funktion eine Zufallsvariable.

2. Sei $\mathcal{S} = \sigma(\mathcal{E})$ für ein $\mathcal{E} \subset 2^S$, dann ist X eine Zufallsvariable, sofern gilt

$$\forall\, C \in \mathcal{E} : X^{-1}(C) \in \mathcal{F}$$

(d.h. es genügt Messbarkeit auf einer Erzeugermenge der σ-Algebra zu prüfen, speziell für $S = \mathbb{R}$ [oder $\overline{\mathbb{R}}$] für Mengen der Form $(-\infty, x]$), denn $\{B \subset S : X^{-1}(B) \in \mathcal{F}\}$ *ist* eine σ-Algebra (Übung), umfasst nach Voraussetzung $\mathcal{E}$.

3. Im Fall $\Omega = \mathbb{R}^m$, $S = \mathbb{R}^d$ (jeweils mit der zugehörigen Borel-σ-Algebra versehen) ist jede stetige Abbildung $f : \Omega \to S$ messbar

(denn $\mathcal{B}(\mathbb{R}^d) = \sigma(\{\text{alle offenen Mengen}\})$
und für stetiges f ist $f^{-1}(\mathcal{O})$ offen,
wenn $\mathcal{O}$ offen ist)

Für $B \subset \{0,1\}$:

$$\mathbb{1}_A^{-1}(B) = \begin{cases} \Omega, & B = \{0,1\} \\ A, & B = \{1\} \\ A^c, & B = \{0\} \\ \emptyset, & B = \emptyset \end{cases}$$

$\cap$

$\mathcal{F}$

Beispiel 1.30 (Indikatorvariable). $A \in \mathcal{F}$, $\mathbf{1}_A : \Omega \to \{0,1\}$,

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

$\mathbf{1}_A$ heißt die Indikatorvariable des Ereignisses A.

Zufallsvariablen sind nicht zuletzt deshalb nützlich für die Modellierung zufälliger Vorgänge, weil man mit ihnen gewissermaßen genauso operieren und „rechnen" kann wie mit anderen Variablen (oder „unbestimmten Größen") in der Mathematik:

Beobachtung 1.31 (Rechnen mit ZVn)**.** Sind $X_1, X_2, \dots$ reelle ZVn, so sind auch

$$(X_1, X_2), X_1 + X_2, X_1 - X_2, X_1 \cdot X_2, X_1 \wedge X_2, X_1 \vee X_2,$$
$$\sup_n X_n, \inf_n X_n, \limsup_{n\to\infty} X_n \text{ und } \liminf_{n\to\infty} X_n$$

($X_1 \wedge X_2$:= Minimum der beiden Werte)

Zufallsvariablen.

Denn: $\{(X_1, X_2) \in (-\infty, x_1] \times (-\infty, x_2]\}$

$= \underbrace{\{X_1 \leq x_1\}}_{\in \mathcal{F}} \cap \underbrace{\{X_2 \leq x_2\}}_{\in \mathcal{F}} \in \mathcal{F}$

die Abbildungen $(x_1, x_2) \mapsto x_1 + x_2$,, $\mapsto x_1 \cdot x_2$
,, $\mapsto x_1 - x_2$ ... etc. sind stetig

$\{\sup_n X_n \leq x\} = \bigcap_{n=1}^{\infty} \underbrace{\{X_n \leq x\}}_{\in \mathcal{F}} \in \mathcal{F}$, analog mit Infimum,

$\limsup_n X_n = \inf_n \sup_{m \geq n} X_m$

Insbesondere können wir das Ereignis

$$\{X_n \text{ konvergiert}\} = \{\limsup_{n\to\infty} X_n \leq \liminf_{n\to\infty} X_n\} \quad \left(= \{(\limsup_{n\to\infty} X_n, \liminf_{n\to\infty} X_n) \in \{(x,y) \in \overline{\mathbb{R}}^2 : x \leq y\}\}\right)$$

sinnvoll betrachten.