Diskrete Zufallsvariable

Wir gehen von einem diskreten W.-raum Ω aus.

- Eine Abbildung

$$X : \Omega \rightarrow \mathbb{R}$$

  heißt *diskrete (numerische) Zufallsvariable* oder kurz *ZV*.

- Der Wertebereich $W_X := X(\Omega)$ ist dann eine diskrete (endliche oder abzählbar unendliche) Menge der Form

$$W_X := \{x_1, x_2, \ldots\} \ .$$

Wertebereich als diskreter W.-raum

- Wahrscheinlichkeiten auf den Elementarereignissen in Ω induzieren Wahrscheinlichkeiten auf den „Elementarereignissen" $x_1, x_2, \ldots$ in W_X:

$$\Pr[X = x_i] := \Pr[X^{-1}(x_i)] = \Pr[\{\omega \in \Omega |\ X(\omega) = x_i\}]$$

- Für „Ereignisse" $E \subseteq W_X$ erhalten wir:

$$\Pr[X \in E] := \Pr[X^{-1}(E)] = \Pr[\{\omega \in \Omega |\ X(\omega) \in E\}$$

Dichte und Verteilung

- Funktion $f_X : \mathbb{R} \to [0,1]$ mit

$$f_X(x) := \Pr[X = x]$$

heißt *diskrete Dichtefunktion* oder kurz *Dichte* von X.

- Funktion $F_X : \mathbb{R} \to [0,1]$ mit

$$F_X(x) := \Pr[X \le x]$$

heißt *diskrete Verteilungsfunktion* oder kurz *Verteilung* von X.

Beispiel 1

Betrachte den W.-raum $\Omega_6^2 = \{1,\ldots,6\} \times \{1,\ldots,6\}$ für zweimaliges Schmeißen eines fairen Würfels und
die Zufallsvariable $S(i,j) := i+j$ (Summe beider Augenzahlen)
mit Wertebereich $W_S = \{2,3,\ldots,11,12\}$:

$$\begin{array}{rcl} S^{-1}(2) & = & \{(1,1)\} \\ S^{-1}(3) & = & \{(1,2),(2,1)\} \\ & \ldots & \\ S^{-1}(7) & = & \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \\ & \ldots & \\ S^{-1}(11) & = & \{(5,6),(6,5)\} \\ S^{-1}(12) & = & \{(6,6)\} \end{array}$$

Es ergibt sich die Dichte

$$f_S(7 \pm i) = \frac{6-i}{36} \quad .$$

Erwartungswert

Der *Erwartungswert* einer Zufallsvariablen X ergibt sich wie folgt:

$$\mathbb{E}[X] := \sum_{x \in W_X} \Pr[X = x] \cdot x = \sum_{x \in W_X} f_X(x) \cdot x$$

Beispiel 1 (fortgesetzt)

Wegen der Symmetrie-Eigenschaft $f_S(7-i) = f_S(7+i)$ gilt $\mathbb{E}[S] = 7$. Ausführlich:

$$
\begin{aligned}
\mathbb{E}[S] &= f_S(7)\cdot 7 + \sum_{i=1}^{5} f_S(7-i)\cdot(7-i) + \sum_{i=1}^{5} f_S(7+i)\cdot(7+i) \\
&= \underbrace{\left(f_S(7)\cdot 7 + \sum_{i=1}^{5}(f_S(7-i) + f_S(7+i)\cdot 7\right)}_{=7} \\
&\quad + \underbrace{\left(\sum_{i=1}^{5}(f_S(7+i)i - f_S(7-i)i)\right)}_{=0} \\
&= 7
\end{aligned}
$$

Eine zweite Darstellung des Erwartungswertes

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_x \Pr[X = x]x \\
&= \sum_x \left(\sum_{\omega : X(\omega) = x} \Pr[\omega] \right) x \\
&= \sum_{(\omega, x) : X(\omega) = x} \Pr[\omega]x \\
&= \sum_\omega \Pr[\omega]X(\omega)
\end{aligned}
$$

Folgerungen

Aus $\mathbb{E}[X] = \sum_\omega \Pr[\omega] X(\omega)$ ergeben sich leicht die folgenden Resultate.

1. Mit der Kurzschreibweise „$X \le Y$“ für „$X(\omega) \le Y(\omega)$ für alle $\omega \in \Omega$“ gilt:

$$X \le Y \implies \mathbb{E}[X] \le \mathbb{E}[Y]$$

2. Der Erwartungswert erfüllt die Linearitätsbedingung

$$\mathbb{E}[a_1 X_1 + a_2 X_2 + \cdots] = a_1 \mathbb{E}[X_1] + a_2 \mathbb{E}[X_2] + \cdots \quad .$$

Die zweite Gleichung ergibt sich dabei wie folgt:

$$\begin{aligned}
\mathbb{E}[a_1 X_1 + a_2 X_2 + \cdots] &= \sum_\omega \Pr[\omega](a_1 X_1(\omega) + a_2 X_2(\omega) + \cdots) \\
&= a_1 \sum_\omega \Pr[\omega] X_1(\omega) + a_2 \sum_\omega \Pr[\omega] X_2(\omega) + \cdots \\
&= a_1 \mathbb{E}[X_1] + a_2 \mathbb{E}[X_2] + \cdots
\end{aligned}$$

Eine dritte Darstellung des Erwartungswertes

Für ZV X mit $W_X \subseteq \mathbb{N}_0$ gilt:

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{i\geq 0} \Pr[X=i]\cdot i \\
&= \sum_{i\geq 1} \Pr[X=i]\cdot i \\
&= \sum_{i\geq 1}\sum_{j=1}^{i} \Pr[X=i] \\
&= \sum_{(i,j):1\leq j\leq i} \Pr[X=i] \\
&= \sum_{j\geq 1}\sum_{i\geq j} \Pr[X=i] \\
&= \sum_{j\geq 1} \Pr[X\geq j]
\end{aligned}
$$

Bedingte Zufallsvariablen

Für eine ZV $X : \Omega \to \mathbb{R}$ und ein Ereignis $A \subseteq \Omega$ bezeichnet $X_{|A}$ die Einschränkung von Abbildung X auf A. Man muss dann bei allen Begriffen (wie Dichte, Verteilung, Erwartungswert, ...) die Wahrscheinlichkeiten $\Pr[\cdot]$ durch die bedingten Wahrscheinlichkeiten $\Pr[\cdot|A]$ ersetzen:

$$
\begin{aligned}
f_{X_{|A}}(x) &:= \Pr[X = x|A] = \Pr[\{\omega : X(\omega) = x\}|A] \\
F_{X_{|A}}(x) &:= \Pr[X \le x|A] = \Pr[\{\omega : X(\omega) \le x\}|A] \\
E[X_{|A}] &:= \sum_{x \in W_X} \Pr[X = x|A]x
\end{aligned}
$$

In Analogie zum Satz von der totalen Wahrscheinlichkeit gilt für jede Zerlegung von Ω in paarweise disjunkte Ereignisse $A_1, \ldots, A_n$ mit jeweils echt positiver Wahrscheinlichkeit:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \Pr[A_i]\mathbb{E}[X_{|A_i}] \ .$$

Zusammengesetzte Zufallsvariablen

Eine ZV $X : \Omega \rightarrow \mathbb{R}$ und eine Abbildung $g : \mathbb{R} \rightarrow \mathbb{R}$ können zu einer neuen ZV

$$Y(\omega) := (g \circ X)(\omega) = g(X(\omega))$$

zusammengesetzt werden. Wie die nachfolgende Rechnung ergibt, gilt

$$\mathbb{E}[Y] = \mathbb{E}[g \circ X] = \sum_{x \in W_X} \Pr[X = x] g(x) \ .$$

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{y \in W_Y} \Pr[Y = y] y \\
&= \sum_{y \in W_Y} y \sum_{x : g(x) = y} \Pr[X = x] \\
&= \sum_{(x,y) : g(x) = y} Pr[X = x] y \\
&= \sum_{x \in W_X} \Pr[X = x] g(x)
\end{aligned}$$

Mehrere Zufallsvariable und Unabhängigkeit

$X_1, \ldots, X_n$ heißen *unabhängig* **gdw** für alle $(x_1, \ldots, x_n) \in W_{X_1} \times \cdots \times W_{X_n}$ gilt:

$$\Pr[X_1 = x_1, \ldots, X_n = x_n] = \Pr[X_1 = x_1] \cdots \Pr[X_n = x_n]$$

Hieraus lässt sich leicht ableiten:

$$\Pr[X_1 \in S_1, \ldots, X_n \in S_n] = \Pr[X_1 \in S_1] \cdots \Pr[X_n \in S_n]$$

Die *gemeinsame Dichte* von $X_1, \ldots, X_n$ ist definiert als

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) := \Pr[X_1 = x_1, \ldots, X_n = x_n]$$

und die *gemeinsame Verteilung* als

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) := \Pr[X_1 \le x_1, \ldots, X_n \le x_n] \ .$$

Für unabhängige Zufallsvariable gilt somit:

$$\begin{aligned} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= F_{X_1}(x_1) \cdots F_{X_n}(x_n) \end{aligned}$$

Beispiele

Beim Skatspiel sind die Zufallsvariablen

$$\begin{array}{rcl} \mathrm{FARBE} & \in & \{\mathrm{Karo,Herz,Pik,Kreuz}\} \\ \mathrm{WERT} & \in & \{7, 8, 9, 10, \mathrm{Bube,Dame,König,As}\} \end{array}$$

unabhängig. Zum Beispiel gilt

$$\Pr[\text{Kreuz Bube}] = \frac{1}{32} = \frac{1}{4} \cdot \frac{1}{8} = \Pr[\text{Kreuz}] \cdot \Pr[\text{Bube}]$$

und wir erhalten die analoge Gleichung für alle anderen Kombinationen von Farbe und Wert. Hingegen sind die Zufallsvariablen

$$\begin{array}{rcl} X & := & \text{Anzahl der Buben auf der Hand} \\ Y & := & \text{Anzahl der Buben im Skat} \end{array}$$

abhängig. Zum Beispiel ist die Kombination $X = 4, Y = 1$ unmöglich (da es nur vier Buben im Spiel gibt), aber jedes der Ereignisse $X = 4$ und $Y = 1$ hat für sich genommen eine echt positive Wahrscheinlichkeit.

Unkorreliertheit

ZV X, Y heißen *unkorreliert* **gdw**

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \ .$$

ZV $X_1, \ldots, X_n$ heißen *unkorreliert* **gdw** für jede Teilmenge $I \subseteq \{1, \ldots, n\}$ gilt:

$$\mathbb{E}\left[\prod_{i \in I} X_i\right] = \prod_{i \in I} \mathbb{E}[X_i] \ .$$

Unabhängigkeit ist stärker als Unkorreliertheit

Satz: Unabhängige ZV sind unkorreliert.

denn: Für unabhängige ZV X, Y gilt

$$
\begin{aligned}
E[X \cdot Y] &= \sum_x \sum_y \Pr[X = x, Y = y] xy \\
&= \sum_x \sum_y \Pr[X = x] \Pr[Y = y] xy \\
&= \left(\sum_x \Pr[X = x] x \right) \cdot \left(\sum_y \Pr[Y = y] y \right) \\
&= \mathbb{E}[X] \cdot \mathbb{E}[Y]
\end{aligned}
$$

und die analoge Rechnung lässt sich für größere Kollektionen von unabhängigen ZV durchführen.

Die Umkehrung des Satzes gilt i.A. nicht, da sich ZV X, Y angeben lassen, die zwar unkorreliert aber nicht unabhängig sind. S. Übung.

Faires Roulette und Risiko

- Bei einem *fairen Spiel* ist der erwartete Gewinn gleich dem Einsatz.
- „Faires Roulette“ sei das „Roulette-Spiel ohne die Null”. (Es ist fair!)
- Wenn wir 1 EUR auf ROT setzen, ist der erwartete Gewinn (bei den vorhandenen Farben ROT und SCHWARZ)
$$\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0 = 1 \ .$$
- Wenn wir 1 EUR auf die Zahl 1 setzen, dann ist der erwartete Gewinn (bei den vorhandenen Zahlen $1, \ldots, 36$)
$$\frac{1}{36} \cdot 36 + \frac{35}{36} \cdot 0 = 1 \ .$$
- Obwohl die Erwartungswerte übereinstimmen, wirkt die zweite Strategie riskanter als die erste: die ZV Gewinn weist eine stärkere mittlere Abweichung vom Erwartungswert auf.

Kovarianz

Die *Kovarianz* der ZV X und Y ist gegeben durch

$$\begin{aligned}\mathrm{Cov}[X,Y] &:= \mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right] \\ &= \mathbb{E}\left[XY-\mathbb{E}[X]Y-\mathbb{E}[Y]X+\mathbb{E}[X]\mathbb{E}[Y]\right] \\ &= \mathbb{E}[XY]-2\mathbb{E}[X]\mathbb{E}[Y]+\mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]\ .\end{aligned}$$

Offensichtlich gilt:

$$\mathrm{Cov}[X,Y]=0 \Leftrightarrow X,Y \text{ sind unkorreliert}$$

Intuition zur Kovarianz

- Große positive Werte der Covarianz zeigen an, dass die ZV X, Y dazu tendieren, *gemeinsam nach oben* und *gemeinsam nach unten* von ihrem jeweiligen Erwartungswert abzuweichen.
- Große negative Werte der Covarianz zeigen an, dass die ZV X dazu tendiert, nach unten von $\mathbb{E}[X]$ abzuweichen, wenn Y nach oben von $\mathbb{E}[Y]$ abweicht (und umgekehrt).
- Kovarianz Null zeigt an, dass die Tendenzen gleich- und gegenläufiger Abweichung vom Erwartungswert sich ausgleichen.

Anwendung: Auf dem Kapitalmarkt sind Portfolios, die negative Kovarianz einbauen, weniger riskant (weil tendenziell Verluste auf einer Seite durch Gewinne auf einer anderen Seite ausgeglichen werden).

Varianz und Standardabweichung

Kovarianz im Spezialfall $Y = X$ führt uns zum Begriff der Varianz:

$$\mathrm{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$\sqrt{\mathrm{Var}[X]}$ heißt auch *Standardabweichung* (oder *Streuung*) von X.

Wir kehren zum Roulettetisch zurück. G_{rot} sei der Gewinn beim Setzen auf ROT und G_1 der Gewinn beim Setzn auf „1“ (Einsatz jeweils 1 EUR). Wir wissen bereits, dass $\mathbb{E}[G_{rot}] = \mathbb{E}[G_1] = 1$. Aus

$$\mathbb{E}[G_{rot}^2] = \frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 0 = 2 \text{ und } \mathbb{E}[G_1^2] = \frac{1}{36} \cdot 36^2 + \frac{35}{36} \cdot 0^2 = 36$$

folgt

$$\mathrm{Var}[G_{rot}] = 2 - 1 = 1 \text{ und } \mathrm{Var}[G_1] = 36 - 1 = 35\ .$$

Rechenregeln zu Varianz und Kovarianz

Für ZV $X, Y, X_1, \ldots, X_n$ und Konstanten $a, b \in \mathbb{R}$ gilt folgendes:

$$\begin{aligned}
\mathrm{Var}[a \cdot X] &= a^2 \cdot \mathrm{Var}[X] \\
\mathrm{Var}[X + b] &= \mathrm{Var}[X] \\
\mathrm{Var}[X + Y] &= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X, Y] \\
\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] &= \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}[X_i, X_j]
\end{aligned}$$

Somit gilt für paarweise unkorrelierte ZV $X_1, \ldots, X_n$

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] \ .$$

Nachweis einer Rechenregel

Wir beschränken uns auf den Nachweis von

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y] \ .$$

Es gilt

$$\mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2+Y^2+2XY] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] \ .$$

Weiterhin gilt

$$\mathbb{E}[X+Y]^2 = (\mathbb{E}[X]+\mathbb{E}[Y])^2 = \mathbb{E}[X]^2 + \mathbb{E}[Y]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] \ .$$

Schließlich erhalten wir

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 \\
&= (\mathbb{E}[X^2] - \mathbb{E}[X]^2) + (\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y] \ .
\end{aligned}$$

Indikatorvariablen

Die *Indikatorvariable* zu einem Ereignis $A \subseteq \Omega$ ist die folgende ZV:

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

Beachte, dass $I_A^2 = I_A$ (wegen $1^2 = 1$ und $0^2 = 0$). Mit $p := \Pr[A]$ gilt:

$$\begin{aligned} \mathbb{E}[I_A] &= \Pr[\bar{A}] \cdot 0 + \Pr[A] \cdot 1 = \Pr[A] = p \\ \mathrm{Var}[I_A] &= \mathbb{E}[I_A^2] - E[I_A]^2 = p - p^2 = p(1-p) \\ I_{A_1} \cdots I_{A_n} &= I_{A_1 \cap \cdots \cap A_n} \\ \mathbb{E}[I_{A_1} \cdots I_{A_n}] &= \mathbb{E}[I_{A_1 \cap \cdots \cap A_n}] = \Pr[A_1 \cap \cdots \cap A_n] \end{aligned}$$

Indikatorvariablen (fortgesetzt)

Weiterhin gilt (s. Übung), dass folgende Aussagen äquivalent sind:

(1) $I_{A_1}, \ldots, I_{A_n}$ sind unkorreliert.

(2) $A_1, \ldots, A_n$ sind unabhängige Ereignisse.

(3) $I_{A_1}, \ldots, I_{A_n}$ sind unabhängige ZV.

Anwendung: Beweis der Siebformel

$$
\begin{aligned}
\Pr[A_1 \cup \cdots \cup A_n] &= 1 - \Pr\left[\overline{A_1 \cup \cdots \cup A_n}\right] \\
&= 1 - \Pr\left[\overline{A_1} \cap \cdots \cap \overline{A_n}\right] \\
&= 1 - \mathbb{E}\left[I_{\overline{A_1} \cap \cdots \cap \overline{A_n}}\right] \\
&= 1 - \mathbb{E}\left[(1 - I_{A_1}) \cdots (1 - I_{A_n})\right] \\
&\overset{*}{=} 1 - \mathbb{E}\left[1 - \sum_{l=1}^{n} (-1)^{l-1} \sum_{1 \le i_1 < \cdots < i_l \le n} I_{A_{i_1}} \cdots I_{A_{i_l}}\right] \\
&= \sum_{l=1}^{n} (-1)^{l-1} \sum_{1 \le i_1 < \cdots < i_l \le n} \mathbb{E}\left[I_{A_{i_1} \cap \cdots \cap A_{i_l}}\right] \\
&= \sum_{l=1}^{n} (-1)^{l-1} \sum_{1 \le i_1 < \cdots < i_l \le n} \Pr[A_{i_1} \cap \cdots \cap A_{i_l}]
\end{aligned}
$$

An der mit „$*$“ gekennzeichneten Stelle wurde einfach ausgeklammert.

Spezielle Wichtige Verteilungen

Bernoulli-Verteilung

Eine Zufallsvariable X mit Wertebereich $W_X = \{0, 1\}$ und Dichte

$$f_X(x) = \begin{cases} p & \text{falls } x = 1 \\ 1 - p & \text{falls } x = 0 \end{cases}$$

heißt *Bernoulli-verteilt mit Erfolgswahrscheinlichkeit* p.

Beispiel: $X = I_A$ für ein Ereignis $A \subseteq \Omega$ mit $\Pr[A] = p$.

Erwartungswert und Varianz sind uns durch das Beispiel der Indikatorvariable bekannt:

$$\mathbb{E}[X] = p \text{ und } \mathrm{Var}[X] = p(1-p)$$

Binomialverteilung

Eine Zufallsvariable X mit Wertebereich $W_X = \{0, 1, \ldots, n\}$ und Dichte

$$f_X(x) = b(x; n, p) := \binom{n}{x} p^x (1-p)^{n-x}$$

heißt *Binomialverteilt bei n unabhängigen Versuchen mit Erfolgswahrscheinlichkeit p.*

Beispiel: $X := X_1 + \cdots + X_n$, wobei $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit Erfolgswahrscheinlichkeit p sind.

Erwartungswert und Varianz lassen sich mit Hilfe dieses Beispiels leicht ermitteln:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np \text{ und } \mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = np(1-p)$$

$b(x; n, p)$ hat eine Maximumstelle bei $x = np$ und fällt zu den Rändern hin „rasch“ ab. (S. Tafelbild bei der Vorlesung.)

Geometrische Verteilung

Eine Zufallsvariable X mit Wertebereich $W_X = \mathbb{N}$ und Dichte

$$f_X(i) = p(1-p)^{i-1}$$

heißt *geometrisch verteilt bei Erfolgswahrscheinlichkeit* p.

Beispiel (Warten auf den Erfolg): Werfe eine Münze mit Wahrscheinlichkeit p für „Kopf“ solange, bis zum ersten Mal „Kopf“ erscheint und setze X auf die Anzahl der hierfür verwendeten Würfe.

Geometrische Verteilung (fortgesetzt)

Setze $q := 1 - p$. Bei der Bestimmung von Erwartungswert und Varianz spielen wir unser Wissen über Potenzreihen und erzeugende Funktionen aus:

$$E[X] = p \cdot \sum_{i \geq 1} iq^{i-1} = p \cdot \frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

Auf ähnliche Weise bestimmen wir

$$E[X^2] = p \cdot \sum_{i \geq 1} i^2 q^{i-1} = p \cdot \frac{1+q}{(1-q)^3} = \frac{p(1+q)}{p^3} = \frac{1+q}{p^2} \quad .$$

Wir erhalten schließlich die Varianz

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2} \quad .$$

Gedächtnislosigkeit der geometrischen Verteilung

Für eine geometrisch verteilte Zufallsvariable gilt:

$$\Pr[X = y + x | X > x] = \Pr[X = y]$$

Dies ergibt sich aus der Interpretation des „Wartens auf den Erfolg":
Wenn wir wissen, dass die ersten x Münzwürfe jedesmal „Zahl" realisieren, dann stimmt die W.-keit, dass wir noch y zusätzliche Münzwürfe bis zum Erscheinen von „Kopf" benötigen, mit $\Pr[X = y]$ überein.

Anwendung: das „Coupon-Collector“ Problem

- Die RUB verkauft Kartons mit T-Shirts, wobei jeder Karton ein T-Shirt und, als freundliche Beigabe, eines von n Bildern mit Motiven des wunderschönen Campus enthält.
- Wieviele T-Shirts muss man im Mittel erwerben, bis man eine vollständige Sammlung aller n Motive besitzt ?

Die Lösung

- Wir teilen die Karton-Käufe in Phasen ein: Phase i endet mit dem Erwerb eines i-ten (von allen bisherigen verschiedenen) Motives der RUB.
- X_i ist die Anzahl der Karton-Käufe in Phase i. Offensichtlich ist X_i geometrisch verteilt bei Erfolgswahrscheinlichket $\frac{n-i+1}{n}$ und hat daher Erwartungswert $\frac{n}{n-i+1}$.
- Die Gesamtanzahl der Karton-Käufe ist gegeben durch ZV $X = \sum_{i=1}^{n} X_i$. Nun ergibt sich die Lösung des Rätsels:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \cdot \sum_{i=1}^{n} \frac{1}{n} = n \cdot H_n$$

Hierbei ist $H_n = 1 + 1/2 + 1/3 + \cdots + 1/n$ die uns bekannte Harmonische Zahl mit $\ln(n) < H_n < 1 + \ln(n)$ für alle $n \geq 2$.

Poisson-Verteilung

Eine Zufallsvariable X mit Wertebereich $W_X = \mathbb{N}_0$ und Dichte

$$f_X(i) = e^{-\lambda}\frac{\lambda^i}{i!}$$

heißt *Poisson-verteilt mit Parameter* λ. Zum Erwartungswert:

$$\mathbb{E}[X] = \sum_{i\geq 1} e^{-\lambda}\frac{\lambda^i}{i!}i = \lambda e^{-\lambda}\sum_{i\geq 1}\frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda}\sum_{i\geq 0}\frac{\lambda^i}{i!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

Analog:

$$\mathbb{E}[X(X-1)] = \sum_{i\geq 2} e^{-\lambda}\frac{\lambda^i}{i!}i(i-1) = \lambda^2 e^{-\lambda}\sum_{i\geq 2}\frac{\lambda^{i-2}}{(i-2)!} = \lambda^2 e^{-\lambda}\sum_{i\geq 0}\frac{\lambda^i}{i!} = \lambda^2$$

Wegen $X^2 = X(X-1) + X$ ergibt sich die Varianz wie folgt:

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Parameter λ repräsentiert also sowohl den Erwartungswert als auch die Varianz.

Poisson-Verteilung als Grenzwert der Binomialverteilung

Wir betrachten eine Binomialverteilung mit Erfolgswahrscheinlichkeit $p = \lambda/n$ und vollziehen den Grenzübergang $n \to \infty$, wobei

- λ konstant bleibt,
- Erfolgswahrscheinlichkeit $p = \lambda/n$ dadurch gegen 0 konvergiert

Intuition: Extrem viele „Bernoulli-Experimente“ mit extrem kleiner Erfolgswahrscheinlichkeit (mit Anwendungen in der Warteschlangentheorie und bei Rechnernetzen).

Der angekündigte Grenzübergang

$$\begin{aligned}
b\left(k; n, \frac{\lambda}{n}\right) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n^{\underline{k}}}{k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \underbrace{\frac{n^{\underline{k}}}{n^k}}_{\to 1}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}}\frac{\lambda^k}{k!} \\
&\approx \underbrace{e^{-\lambda}\frac{\lambda^k}{k!}}_{\text{Dichte der Poisson-Verteilung}}
\end{aligned}$$

Analyse der Abweichung vom Erwartungsswert

Die Markov'sche Ungleichung

Satz: Für eine ZV $X \geq 0$ und für alle $t > 0$ gilt die Ungleichung

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

oder äquivalent dazu

$$\Pr[X \geq t \cdot \mathbb{E}[X]] \leq \frac{1}{t} \ .$$

Die Wahrscheinlichkeit für eine nicht-negativen ZV, ihren Erwartungswert um den Faktor t (oder stärker) zu überschreiten, ist also durch $1/t$ beschränkt.

Beweis: Wie wir wissen:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X|X \geq t] \cdot \Pr[X \geq t] + \mathbb{E}[X|X < t] \Pr[X < t] \\ &\geq t \cdot \Pr[X \geq t] + 0\end{aligned}$$

Auflösen der Ungleichung nach $\Pr[X \geq t]$ liefert die Behauptung.

Die Chebychev'sche Ungleichung

Satz: Für eine ZV X und für alle $t > 0$ gilt die Ungleichung

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}$$

oder äquivalent dazu

$$\Pr[|X - \mathbb{E}[X]| \geq t \cdot \sqrt{\mathrm{Var}[X]}] \leq \frac{1}{t^2} \ .$$

Die Wahrscheinlichkeit, dass die absolute Abweichung einer ZV von ihrem Erwartungswert das t-fache ihrer Standardabweichung beträgt (oder mehr), ist also durch $1/t^2$ beschränkt.

Beweis: Wir wenden die Markov'sche Ungleichung auf die ZV $(X - \mathbb{E}[X])^2$ an:

$$\begin{aligned} \Pr[|X - \mathbb{E}[X]| \geq t] &= \Pr[(X - \mathbb{E}[X])^2 \geq t^2] \\ &\leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{\mathrm{Var}[X]}{t^2} \end{aligned}$$

Chernov-Schranken (additive Form)

Wir setzen folgendes voraus:

- $X_1, \ldots, X_n$ sind unabhängige ZV, die jeweils Bernoulli-verteilt mit Erfolgswahrscheinlichkeit p sind (n unabhängige „Bernoulli-Experimente“).
- $Z := \frac{1}{n}(X_1 + \cdots + X_n)$ ist der resultierende „empirische Mittelwert“.

Dann gilt für alle $0 \leq \delta \leq 1$:

$$\begin{aligned} \Pr[Z \geq p + \delta] &\leq e^{-2\delta^2 n} \\ \Pr[Z \leq p - \delta] &\leq e^{-2\delta^2 n} \\ \Pr[|Z - p| \geq \delta] &\leq 2e^{-2\delta^2 n} \end{aligned}$$

Da $e^{-2\delta^2 n}$ mit wachsendem n in mit exponentieller Geschwindigkeit gegen Null konvergiert, sind empirische Schätzungen für die Erfolgswahrscheinlichkeit p schon bei moderaten Stichprobengrößen n mit „überwältigender W.-keit“ sehr genau.

Chernov-Schranken (multiplikative Form)

Unter denselben Voraussetzungen wie eben gilt:

$$\begin{aligned}\Pr[Z \geq (1+\delta)p] &\leq e^{-np\delta^2/3}\\ \Pr[Z \leq (1-\delta)p] &\leq e^{-np\delta^2/2}\end{aligned}$$

Hoeffding-Schranken

Die Hoeffding-Schranken sind Verallgemeinerungen der Chernov-Schranken auf beschränkte ZV (die nicht notwendig Bernoulli-verteilt sind). Wir beschränken uns hier auf die additive Form.

Wir setzen folgendes voraus:

- $X_1, \ldots, X_n$ sind unabhängige und gleichverteilte ZV mit Werten im Intervall $[a, b]$ und Erwartungswert μ.
- $Z := \frac{1}{n}(X_1 + \cdots + X_n)$ ist der resultierende „empirische Mittelwert“ (bei n unabhängigen Versuchen).

Dann gilt für alle $0 \leq \delta \leq 1$:

$$\begin{aligned}
\Pr[Z \geq \mu + \delta] &\leq e^{\frac{-2\delta^2 n}{(b-a)^2}} \\
\Pr[Z \leq \mu - \delta] &\leq e^{\frac{-2\delta^2 n}{(b-a)^2}} \\
\Pr[|Z - \mu| \geq \delta] &\leq 2e^{\frac{-2\delta^2 n}{(b-a)^2}}
\end{aligned}$$