

Wahrscheinlichkeitstheorie – Diskrete Verteilungen

Holger Jäkel

Communications Engineering Lab (CEL)

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

4 Diskrete Verteilungen

- Wiederholung
- Zufallsvariablen
- Kenngrößen von Zufallsvariablen
- Wichtige Verteilungen
- Lernziele
- Literatur

- Folgende Diskussionen erfolgen gemäß¹

[JW02]: F. Jondral, A. Wiesler, *Wahrscheinlichkeitsrechnung und stochastische Prozesse*, Teubner, 2002

¹ Von dort entstammt die Struktur der Folien, die Formeln und die meisten Bilder.

4 Diskrete Verteilungen

- Wiederholung
- Zufallsvariablen
- Kenngrößen von Zufallsvariablen
- Wichtige Verteilungen
- Lernziele
- Literatur

■ Erinnerung:

- Ein *Wahrscheinlichkeitsraum* ist ein Tupel $(\Omega, \mathcal{B}, P)$. Hierbei ist
 - Ω der Ergebnisraum,
 - $\mathcal{B}$ eine σ -Algebra auf Ω und
 - die *Wahrscheinlichkeit* eine Abbildung $P : \mathcal{B} \rightarrow \mathbb{R}$ mit:

$$P(A) \geq 0, A \in \mathcal{B}$$

$$P(\Omega) = 1$$

$$P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n), \quad A_n \in \mathcal{B}$$

Der Wert $P(A)$, $A \in \mathcal{B}$, ist die *Wahrscheinlichkeit des Ereignisses* $A \in \mathcal{B}$.

- Ist Ω endlich, so kann stets $\mathcal{B} = \mathcal{P}(\Omega)$ gewählt werden.
- In Laplace'schen Wahrscheinlichkeitsräumen sind zudem alle Elementarereignisse gleichwahrscheinlich; es folgt $P(A) = |A|/|\Omega|$.



4 Diskrete Verteilungen

- Wiederholung
- Zufallsvariablen
- Kenngrößen von Zufallsvariablen
- Wichtige Verteilungen
- Lernziele
- Literatur

- **Grundidee:** Konstruktion eines Wahrscheinlichkeitsraums $(\Omega, \mathcal{B}, P)$ sehr abstrakt; praktisch beobachtet man oft reelle Zahlen²
- **Aufgabe:** Überführen des Wahrscheinlichkeitsbegriffes von Ω auf $\mathbb{R}$
- **Methode:**
 - Verwendung einer Abbildung, im Weiteren *Zufallsvariable* genannt:

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

- Charakterisierung der Wahrscheinlichkeit in $\mathbb{R}$

²Der Fall endlich vieler diskreter Werte ergibt sich als Spezialfall.

- **Frage:** Wieso eine Zufallsvariable einführen? Wieso nicht direkt die auftretenden Werte mit entsprechenden Wahrscheinlichkeiten verwenden?
- **Antwort:** Zitat aus [Sch15, S. 89]:

„[...] Eine Zufallsvariable dient dazu, gewisse Teilaspekte eines Zufallsexperiments [...] hervorzuheben und unwichtige Teilaspekte auszublenden.“



■ Beispiel:

- **Fragestellung:** Es werden 5 Würfel geworfen; deren Augensumme sei, z.B. für ein Gesellschaftsspiel, relevant.
- **Lösung 1:** Summe aus 5 Würfeln $\implies \tilde{\Omega} = \{5, 4, \dots, 30\}$;
Wahrscheinlichkeiten $P(\tilde{\omega})$ durch Abzählen der Möglichkeiten
- **Lösung 2:** Verwende ein Laplace-Modell mit $\Omega = \{1, \dots, 6\}^5$ und gleichwahrscheinlichen Tupeln; Augensumme ist dann „nur noch“ eine Abbildung

- **Erinnerung:** Zufallsvariablen sind Abbildungen $X : \Omega \rightarrow \mathbb{R}$ und „übertragen Wahrscheinlichkeit“ von Ω auf $\mathbb{R}$
- **Problem:** „Wahrscheinlichkeitsmessung“ nur in $(\Omega, \mathcal{B}, P)$ möglich $\implies$ für die „interessanten“ Ereignisse $A \subset \mathbb{R}$, $A \in \mathcal{B}_{\mathbb{R}}$, muss $P(X^{-1}(A))$ berechnet werden können, also muss gelten

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{B}, \quad A \in \mathcal{B}_{\mathbb{R}}$$

Definition

Eine *reellwertige Zufallsvariable* auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{B}, P)$ ist eine Abbildung

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

die jedem Elementarereignis eine reelle Zahl zuordnet und für die gilt:³

$$X^{-1}((-\infty, a]) \in \mathcal{B}$$

Die *Verteilung* einer Zufallsvariablen beschreibt die Verteilung der Wahrscheinlichkeit über $\mathbb{R}$:

$$P_X(A) = P(X^{-1}(A)), \quad A \in \mathcal{B}_{\mathbb{R}}$$

³Die Notwendigkeit dieser Forderung wird im nächsten Kapitel deutlich werden.



■ Beispiel: (Würfelwurf)

- Wähle $\Omega = \{1, 2, \dots, 6\}$ und⁴

$$\mathcal{B} = \sigma(\{\{1, 2\}\}) = \{\emptyset, \{1, 2\}, \{3, 4, 5, 6\}, \Omega\}$$

- Die Augenzahl wird über die Abbildung⁵ $X : \Omega \rightarrow \mathbb{R}, i \mapsto i$ berechnet.
- In diesem Fall folgt:

$$X^{-1}((-\infty, 3]) = \{1, 2, 3\} \notin \mathcal{B}$$

Die Wahrscheinlichkeit $P(X \leq 3) = P(\{1, 2, 3\})$ kann nicht berechnet werden, da $\{X \leq 3\}$ kein zulässiges Ereignis ist.

⁴Vergleiche das Beispiel in Kapitel 2.

⁵Da die Zufallsvariable hier „nichts tut“ ist diese Konstruktion eigentlich unnötig. Zur Verdeutlichung der Aussage ist es dennoch hilfreich.



■ Ergänzungen:

- Das Problem in obigem Beispiel ist, dass $\mathcal{B} = \sigma(\{\{1, 2\}\})$ „zu klein ist“. Wählt man in vorigem Beispiel $\mathcal{B} = \mathcal{P}(\Omega)$, so ist das Urbild ein zulässiges Ereignis, $\{X \leq 3\} = \{1, 2, 3\} \in \mathcal{B}$, und es folgt:

$$P_X((-\infty, 3]) = P(X^{-1}((-\infty, 3])) = P(\{1, 2, 3\}) = \frac{3}{6} = 0.5$$

- Die Mengen $(-\infty, a]$, $a \in \mathbb{R}$, sind „ziemlich viele“, insbesondere relativ zu Ω des obigen Beispiels. Allerdings ist:

$$P_X((-\infty, 3.3]) = P(X^{-1}((-\infty, 3.3])) = P(\{1, 2, 3\}) = \frac{3}{6} = 0.5$$

Es „geht also trotz $|\mathcal{B}| \ll |\{(-\infty, a], a \in \mathbb{R}\}|$ gut“, weil oft⁶ die Gleichheit $X^{-1}((-\infty, a]) = X^{-1}((-\infty, b])$ gilt.

⁶Übung: Überlegen Sie sich, wann das zutrifft.



Definition

Eine reellwertige Zufallsvariablen X ist *diskret*, falls ihr Wertebereich $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ endlich oder abzählbar ist. Der Wertebereich kann damit durch

$$X(\Omega) = \{x_1, x_2, \dots\}$$

beschrieben werden.

Die Verteilung diskreter Zufallsvariablen ist durch die Einzelwahrscheinlichkeiten

$$p_n = P(X = x_n)$$

vollständig beschrieben.

Definition

Die *Verteilungsfunktion* einer diskreten, reellwertigen Zufallsvariablen ist definiert⁷ durch:

$$F_X(x) = P(X \leq x)$$

Die Verteilungsfunktion diskreter Zufallsvariablen ergibt sich zu:

$$F_X(x) = \sum_{n: x_n \leq x} P(X = x_n)$$

⁷Je nach Autor wird die Verteilungsfunktion auch als $\tilde{F}(x) = P(X < x)$ definiert.

■ Bemerkungen/Beispiele:⁸

- Ist der Wahrscheinlichkeitsraum Ω endlich oder abzählbar, so ist jedes reellwertige X diskret. Für das Würfelbeispiel von Folie 12 folgt etwa:

$$X(\Omega) = \{1, 2, 3, 4, 5, 6\}$$

- Aufgrund der Definition von Zufallsvariablen kann die Verteilungsfunktion stets berechnet werden.

⁸Übung: Begründen Sie dies.



■ Beispiel:⁹

- **Szenario:** Bei einem zweifachen Würfelwurf unterscheidbarer Würfel interessiere die Augensumme
- **Modell:** Laplace-Experiment auf¹⁰

$$\Omega := \{(i, j) : 1 \leq i, j \leq 6\}, \quad \mathcal{B} = \mathcal{P}(\Omega)$$

- Die Beschreibung der Summe als Zufallsvariable und ihre Verteilung lauten:

$$X : \begin{cases} \Omega \rightarrow \mathbb{R} \\ \omega = (i, j) \mapsto i + j \end{cases}$$

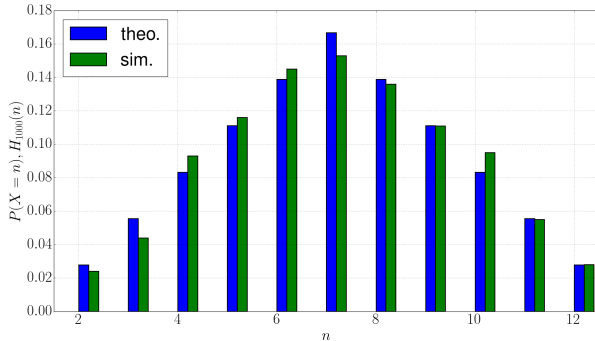
k	2	3	4	5	6	7	8	9	10	11	12
$P(X = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

⁹**Datei:** dice_sum.ipynb

¹⁰**Frage/Übung:** Wieso kann $\mathcal{P}(\Omega)$ verwendet werden?

■ Beispiel:¹¹(ctd.)

- Graphische Darstellung der Verteilung; Simulation von 1 000 Werten



¹¹ Datei: dice_sum.ipynb



■ Beispiel:¹² (ctd.)

- Aus der Verteilung ergibt sich die Verteilungsfunktion...

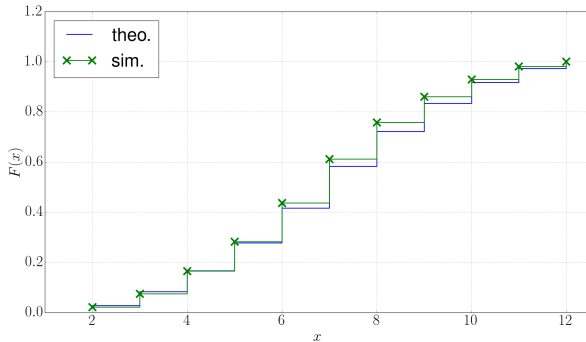
x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

¹²Datei: dice_sum.ipynb



■ Beispiel:¹³ (ctd.)

- Graphische Darstellung der Verteilungsfunktion bei zweifachem Würfelwurf;
Simulation von 1 000 Werten



¹³ Datei: dice_sum.ipynb

- **Beachte:** Je größer der Sprung von $F(x)$ an einer Stelle ist, desto wahrscheinlicher ist der jeweilige Wert.
- **Beispiel:** Demnach sind bei der Summe zweier Würfel 6, 7, 8 am wahrscheinlichsten, was auch durch die Verteilung bestätigt wird.

■ **Beispiel:**¹⁴ („Rauskommen“ bei Mensch ärgere Dich nicht)

- Es wird so lange ein Würfel geworfen bis eine 6 auftritt.
- Die Zufallsvariable X beschreibt die Anzahl der Würfe bis zu ersten 6.
- Der Wertebereich ist $X(\Omega) = \mathbb{N}$.
- Es ergibt sich mit $p = 1/6$:

$$P(X = k) = (1 - p)^{k-1} \cancel{p}^k p$$

¹⁴... für eine Zufallsvariable mit abzählbar unendlichem Wertebereich...

■ Bemerkung:

- Im Beispiel „Augensumme bei zweifachem Würfelwurf“ könnte auch direkt

$$\tilde{\Omega} = \{2, 3, \dots, 12\}$$

mit entsprechender Verteilung $\tilde{P}$ verwendet werden. Dann wäre die Verwendung einer Zufallsvariablen unnötig.

Aber: Durch das Vorgehen wie im Beispiel entsteht für (Ω, P) ein Laplace-Experiment, das „nur noch durch X abgebildet wird“. Dieses Vorgehen ist des öfteren einfacher als die direkte Modellierung mit nicht Laplace'schen Wahrscheinlichkeiten (→ vgl. `dice_sum.ipynb`).

4 Diskrete Verteilungen

- Wiederholung
- Zufallsvariablen
- Kenngößen von Zufallsvariablen**
 - Momente einer Zufallsvariablen
 - Charakteristische Funktion
- Wichtige Verteilungen
- Lernziele
- Literatur

- **Beobachtung I:** Betrachtung der Verteilung einer Zufallsvariablen

$$P_X(x_n) = P(X = x_n) = P(\{\omega \in \Omega : X(\omega) = x_n\})$$

bietet vollständige Information

- **Beobachtung II:** Charakterisierung der kompletten Verteilung oft „unnötig kompliziert“; evtl. genügt weniger Information, die aber leichter zu erfassen ist
- **Idee:** Beschränken auf charakteristische Werte der Verteilung wie etwa
 - Durchschnittliche Lebensdauer
 - Streuung um Mittelwerte

Definition

Der *Erwartungswert* oder *Mittelwert* einer diskreten Zufallsvariablen X mit Wertebereich $X(\Omega) = \{x_1, x_2, \dots\}$ ist definiert durch

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{n=1}^{\infty} x_n P(X = x_n)$$

falls die Summe existiert.



■ **Bemerkung:** (Bezug auf Beispiel der Summe aus 5 Würfeln)

- Bei der Berechnung des Erwartungswerts erkennt man die „Stärke“ der Verwendung von Zufallsvariablen...

■ **Lösung 1:**

- Verwendung von $\tilde{\Omega} = \{5, \dots, 30\}$
- $X(\omega)$: unnötig/nicht vorhanden
- $P(\tilde{\omega})$: schwierig („Rückwärtsrechnung“; „welche Kombinationen führen auf Summe $\tilde{\omega}$?“)

■ **Lösung 2:**

- Verwendung von $\Omega = \{1, \dots, 6\}^5$ mit $\omega = (a_1, \dots, a_5) \in \Omega$ und $P(\omega) = 1/|\Omega|$
- $P(\omega)$: sehr einfach
- $X(\omega)$: relativ einfach („Vorwärtsrechnung“)

Definition

Der *Erwartungswert* oder *Mittelwert* einer diskreten Zufallsvariablen X mit Wertebereich $X(\Omega) = \{x_1, x_2, \dots\}$ bzgl. der Funktion $g(\cdot)$ ist definiert durch

$$E(g(X)) = \sum_{\omega \in \Omega} g(X(\omega))P(\omega) = \sum_{n=1}^{\infty} g(x_n)P(X = x_n)$$

falls die Summe existiert.



Definition

Falls die entsprechenden Summen existieren, so lautet das *k-te Moment*

$$E(X^k) = \sum_{\omega \in \Omega} X^k(\omega) P(\omega) = \sum_{n=1}^{\infty} x_n^k P(X = x_n)$$

und das *k-te zentrale Moment*:

$$E((X - E(X))^k) = \sum_{\omega \in \Omega} (X - E(X))^k(\omega) P(\omega) = \sum_{n=1}^{\infty} (x_n - E(X))^k P(X = x_n)$$



Definition

Das zweite zentrale Moment

$$D^2(X) := \text{var}(X) := E((X - E(X))^2) = E(X^2) - (E(X))^2$$

heißt *Varianz* der Verteilung.

Die *Standardabweichung*¹⁵ ist die Wurzel der Varianz:

$$D(X) := \sqrt{E((X - E(X))^2)}$$

¹⁵Der Begriff „dispersio“ (Lat.) bzw. Dispersion motiviert die Verwendung des Bezeichners $D(\cdot)$. [Ren62]

- **Eigenschaften:** Sind X, Y Zufallsvariablen und $a, b \in \mathbb{R}$ Konstanten, so gilt:¹⁶

$$E(X + b) = E(X) + b$$

$$E(X + Y) = E(X) + E(Y)$$

$$E(aX) = aE(X)$$

$$D^2(aX) = a^2 D^2(X)$$

$$D^2(X + b) = D^2(X)$$

¹⁶**Nachweis:** Entweder selbst als **Übung** nachrechnen; alternativ Nachweis in der Übung abwarten.

Definition

Eine Zufallsvariable X heißt *standardisiert*, falls

$$E(X) = 0$$

$$D^2(X) = 1.$$

Satz

Ist $E(X) = m < \infty$ und $D^2(X) = \sigma^2 < \infty$, so ist

$$Y = \frac{X - m}{\sigma}$$

standardisiert.

■ **Nachweis:** Man rechnet:

$$E(Y) = E\left(\frac{X - m}{\sigma}\right) = \frac{E(X - m)}{\sigma} = 0$$

$$D^2(Y) = D^2\left(\frac{X - m}{\sigma}\right) = \frac{1}{\sigma^2} D^2(X - m) = \frac{1}{\sigma^2} \sigma^2 = 1$$



■ Beispiele:

- Der Würfelwurf hat den Erwartungswert:¹⁷

$$E(X) = \sum_{n=1}^6 \frac{1}{6}n = \frac{1}{6} \sum_{n=1}^6 n = \frac{7 \cdot 6}{6 \cdot 2} = 3.5$$

- Der Würfelwurf hat die Varianz:

$$D^2(X) = \sum_{n=1}^6 (n - 3.5)^2 \frac{1}{6} = \frac{35}{12}$$

- **Übung:** Verwenden Sie die Gewinnquoten der letzten Woche, um den Erwartungswert beim Lotto zu berechnen.¹⁸

¹⁷Dieses Beispiel zeigt, dass der Erwartungswert nicht notwendig ein gültiges Ergebnis ist.

¹⁸**Hinweis:** Erinnern Sie sich an die hypergeometrische Verteilung in Kap. 2 bei der Diskussion der Laplace-Wahrscheinlichkeiten.

Definition

- Ein Wert x_p , für den

$$P(X < x_p) \leq p, \quad P(X > x_p) \leq 1 - p$$

gilt, heißt *p*-tes Quantil. Das $p = \frac{1}{2}$ -Quantil heißt *Median*.



■ Bemerkungen:

- Anschaulich liegt bei einem p -Quantil höchstens p Wahrscheinlichkeit „links davon“ und höchstens $1 - p$ der Wahrscheinlichkeit „rechts davon“.
- Der Median ist robuster gegen Ausreißer, weswegen er in Statistiken¹⁹ häufig eingesetzt wird [Hen13].

Enthält die Beobachtung Ausreißer nach oben, so gilt $x_{\frac{1}{2}} \leq E(X)$, bei Ausreißern nach unten gilt $x_{\frac{1}{2}} \geq E(X)$.

- Quantile sind nicht unbedingt eindeutig.²⁰
Dies kann dadurch vermieden werden, dass das p -Quantil als

$$\inf\{x \in \mathbb{R} : P(X \leq x) \geq p\}$$

definiert wird [Hen13]. Beachten Sie, dass die Literatur hier uneinheitlich ist.

¹⁹Übung: Finden Sie ein Beispiel, in dem der Median aussagekräftiger ist.

²⁰Übung: Finden Sie ein Gegenbeispiel.

■ Beispiele:

- Die Studiendauer an einer Hochschule werde über eine Zufallsvariable $X \in \mathbb{N}$ beschrieben.²¹ Machen $N = 10$ Studierende ihren Abschluss in

7, 8, 9, 9, 10, 10, 11, 11, 12, 51

Semestern, so ist bei Annahme einer Gleichverteilung der Erwartungswert $E(X) = 13.8$ und der Median $x_{\frac{1}{2}} = 10$ bzw. jeder Wert zwischen 10 und 11.

- An einer zweiten Hochschule werden die Abschlüsse nach

2, 7, 8, 9, 9, 10, 10, 11, 11, 12

erreicht. Hier ist: $E(X) = 8.9$ und $x_{\frac{1}{2}} = 9$ bzw. jeder Wert zwischen 9 und 10.

²¹ ... die Studiendauern seien hier nicht durch die PO begrenzt...

Definition

Die *charakteristische Funktion* einer Zufallsvariablen X ist²² definiert als

$$\varphi(s) = \varphi_X(s) := E(e^{jsX}), \quad s \in \mathbb{R}.$$

Es folgt:

$$\varphi(s) = \sum_{\omega \in \Omega} e^{jsX(\omega)} P(\omega) = \sum_{n=1}^{\infty} e^{jsx_n} P(X = x_n)$$

²² ... im Falle der Existenz...

Satz

Falls das k -te Moment der Zufallsvariablen X existiert, so kann es gemäß

$$E(X^k) = \frac{\varphi_X^{(k)}(0)}{j^k}, \quad k = 1, 2, \dots$$

aus der charakteristischen Funktion von X berechnet werden.²³

■ **Nachweis:** Tafel; Ansatz: Ableiten von $\varphi(s)$ ■

²³**Notation:** $\varphi^{(k)}(s) := \frac{d^k}{ds^k} \varphi(s)$

- **Frage:** Welche Umformungen ergeben sich für die charakteristische Funktion diskreter Zufallsvariablen, falls die diskrete Zufallsvariable nur Werte $0, 1, 2, \dots$ annimmt?
- **Antwort:** Für die charakteristische Funktion ergibt sich:

$$\varphi(s) = \sum_{n=0}^{\infty} e^{jsn} P(X = n) = \sum_{n=0}^{\infty} P(X = n) (e^{js})^n$$



Definition

Nimmt die diskrete Zufallsvariable X nur Werte in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ an, so lautet im Falle der Existenz deren *erzeugende Funktion*:

$$\psi(z) = \psi_X(z) := E(z^X) = \sum_{n=0}^{\infty} P(X = n)z^n, \quad z \in \mathbb{C}$$

Satz

Die Verteilung von X ergibt sich bei Existenz aus den Ableitungen der erzeugenden Funktion:

$$P(X = n) = \frac{\psi_X^{(n)}(0)}{\cancel{j^n} n!}, \quad n = 1, 2, \dots$$

■ **Nachweis:** Rechnung wie bei charakteristischer Funktion ■

■ Bemerkungen:

- Wegen

$$|\psi(1)| = \left| \sum_{n=0}^{\infty} P(X = n) \right| \leq \sum_{n=0}^{\infty} P(X = n) = 1$$

existiert die erzeugende Funktion mindestens für $|z| \leq 1$ und ist dort absolut und gleichmäßig konvergent sowie stetig.

- Da eine Funktion nur auf genau eine Weise als Potenzreihe dargestellt werden kann, sind zwei Verteilungen genau dann identisch, wenn ihre erzeugenden Funktionen identisch sind. [Hen13]

4 Diskrete Verteilungen

- Wiederholung
- Zufallsvariablen
- Kenngrößen von Zufallsvariablen
- **Wichtige Verteilungen**
 - Die Zweipunktverteilung
 - Die Binomialverteilung
 - Die Polynomialverteilung
 - Die Poissonverteilung
 - Die Hypergeometrische Verteilung
- Lernziele
- Literatur

- **Motivation:** Einige Verteilungen sind in der Anwendung der Wahrscheinlichkeitstheorie von großer Bedeutung, weil sie sich als Modelle für bestimmte Phänomene bewährt haben.
- **Beispiele:**²⁴
 - Modellieren von Bit- und Symbolfehlern (unabhängige Fehler oder Bündelfehler)
 - Verteilung von Fehlern nach/trotz Codierung
 - Modellieren der Anzahl eintreffender Pakete in nachrichtentechnischen Systemen

²⁴Es sei die Egozentrik verziehen, dass die Nachrichtentechnik in dieser Auflistung dominiert. Aufgrund des eigenen Arbeitsgebiets liegen diese Beispiele nahe.

■ Modell:

- Zufallsvariable X kann zwei mögliche Werte x_1, x_2 mit $x_2 < x_1$ annehmen
- Der Beobachter sieht entweder $A = \{X = x_1\}$ oder $\bar{A} = \{X = x_2\}$

■ Anwendung/Beispiele:

- Münzwurf
- Wert eines zufälligen Bits
- Radar-Detektion

■ Kenngrößen der Verteilung:²⁵

$$P(A) = p, \quad P(\overline{A}) = 1 - p$$

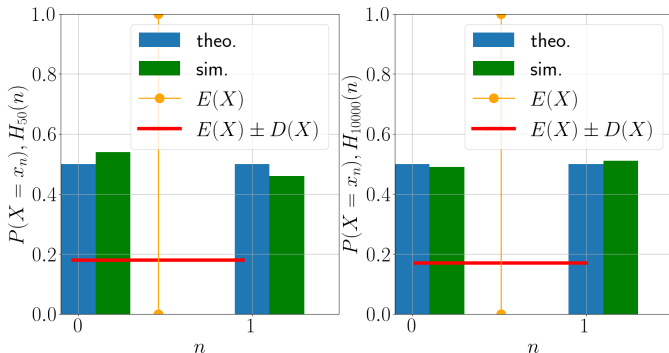
$$E(X) = x_1 p + x_2 (1 - p)$$

$$D^2(X) = (x_1 - x_2)^2 p (1 - p)$$

$$\varphi(s) = e^{jsx_2} + p (e^{jsx_1} - e^{jsx_2})$$

²⁵Erinnerung: $x_2 < x_1$

- **Beispiel:**²⁶ Simulation von $N_1 = 50$ bzw. $N_2 = 10\,000$ Würfeln eines fairen Bits²⁷



²⁶ **Datei:** two_point.ipynb

²⁷ Damit ist: $x_2 = 0, x_1 = 1 \implies E(X) = 0.5, D^2(X) = 0.25$

■ Ergänzung:

- Das (Nicht-)Eintreten von Ereignissen spielt – insbesondere in Produktexperimenten – eine wichtige Rolle
- Mathematisch kann dies durch eine Zufallsvariable („Funktion“) modelliert werden, die bei gegebenem Ergebnis eines Zufallsexperiments prüft, ob das Ereignis eingetreten ist...

Definition

Die *Indikatorfunktion* $I_A : \Omega \rightarrow \mathbb{R}$ zeigt an, ob ein bestimmtes Ereignis $A \in \mathcal{B}$ eintritt. Sie ist definiert durch:

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

Es ist:

$$P(I_A = 1) = P(A) =: p$$

$$E(I_A) = p$$

$$D^2(I_A) = p(1 - p)$$

■ Modell: (*Bernoulli'sches Versuchsschema*)

- Durchführen von N unabhängigen Versuchen mit Zweipunktverteilung auf den Ereignissen $A = \{X = x_1\}$, $\bar{A} = \{X = x_2\}$ mit $P(A) = p$
- Wahrscheinlichkeitsraum:

$$\Omega^N = \Omega \times \cdots \times \Omega$$

$$\omega = (a_1, \dots, a_N), a_i \in \Omega$$

- X_N beschreibt die Anzahl der Versuche mit Ergebnis A („Treffer“):

$$X_N = \sum_{i=1}^N I_A(a_i) = |\{i : \text{Versuch } i \text{ ergibt } A\}| \in \{0, 1, \dots, N\}$$

■ Anwendung/Beispiele:

- Anzahl fehlerhafter Bits bei Übertragung von N Bits
- N -faches Würfeln und Zählen der 6-er

■ Kenngrößen der Verteilung:²⁸

$$P(X_N = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

$$E(X_N) = Np$$

$$D^2(X_N) = Np(1 - p)$$

$$\varphi(s) = (1 + p(e^{js} - 1))^N$$

²⁸ **Übung:** Erklären Sie den Erwartungswert anschaulich.

■ Bemerkungen:

- Eine Binomialverteilung wird durch die Parameter N und p beschrieben
- Eine Binomialverteilung zählt die absolute Häufigkeit des Ereignisses A :

$$X_N = h_N(A)$$

Teilt man durch die Anzahl der Versuche, so gilt im Erwartungswert:

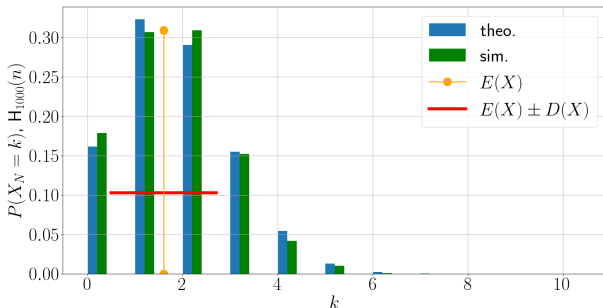
$$E(H_N(A)) = E\left(\frac{h_N(A)}{N}\right) = \frac{1}{N}E(X_N) = p$$

$$D^2(H_N(A)) = D^2\left(\frac{h_N(A)}{N}\right) = \frac{1}{N^2}D^2(X_N) = \frac{1}{N}p(1-p)$$

Im Mittel tritt das Ereignis also mit relativer Anzahl p auf, die Varianz ist invers proportional zu N .

- **Beispiel:**²⁹ Anzahl der 6-en bei 10-maligem Würfeln.

$$P(X_{10} = k) = \binom{10}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k}$$



²⁹ **Datei:** dice_num_6.ipynb



- **Beispiel:** Bei der Übertragung eines Datenpakets mit $N = 100$ Bits wird jedes Bit unabhängig von den anderen mit Wahrscheinlichkeit $p = 10^{-3}$ verfälscht. Berechnen Sie
 - die Wahrscheinlichkeit, dass das Paket korrekt übertragen wird.
 - die Wahrscheinlichkeit, dass mehr als 4 Fehler auftreten.³⁰
- **Lösung:** Man rechnet:

$$P(X_N = 0) = \binom{100}{0} p^0 (1-p)^{N-0} = 0.999^{100} \approx 0.9047$$

$$\begin{aligned} P(X_N > 4) &= \sum_{k=5}^{100} \binom{100}{k} p^k (1-p)^{N-k} \\ &= 1 - \sum_{k=0}^4 \binom{100}{k} p^k (1-p)^{N-k} \approx 7 \cdot 10^{-8} \end{aligned}$$

³⁰ Relevante Fragestellung, falls im Empfänger bis zu 4 Fehler korrigiert werden können.

■ Modell:

- Durchführen von N unabhängigen Versuchen mit je $r + 1$ möglichen Ergebnissen $A_1, \dots, A_{r+1}$ mit $P(A_j) = p_j, j = 1, \dots, r + 1$
- Wahrscheinlichkeitsraum:

$$\Omega^N = \Omega \times \dots \times \Omega$$

$$\omega = (a_1, \dots, a_N), a_i \in \Omega$$

- X_j beschreibt die Anzahl der Versuche mit Ergebnis A_j („Treffer“):

$$X_j = \sum_{i=1}^N I_{A_j}(a_i) = |\{i : \text{Versuch } i \text{ ergibt } A_j\}| \in \{0, 1, \dots, N\}$$

■ Bemerkung:

- Beschreibt eine Verallgemeinerung der Binomialverteilung
- Beachte, dass stets $p_1 + \dots + p_{r+1} = 1$ und $X_1 + \dots + X_{r+1} = N$. Die Angabe von p_{r+1} und X_{r+1} ist somit redundant.³¹
- Jedes X_j ist binomialverteilt mit $P(A_j) = p_j$ und $P(\overline{A_j}) = \sum_{i \neq j} p_i = 1 - p_j$

³¹ In [JW02] wird die „umständliche“ obere Grenze $r + 1$ gewählt, um zu zeigen, dass nur r Werte der Wahrscheinlichkeit wählbar sind.

- **Verteilung:** Für $K_1 + \dots + K_{r+1} = N$ folgt:

$$\begin{aligned} P(X_1 = K_1, X_2 = K_2, \dots, X_{r+1} = K_{r+1}) \\ = \frac{N!}{K_1! \cdot K_2! \cdot \dots \cdot K_{r+1}!} \cdot p_1^{K_1} \cdot p_2^{K_2} \cdot \dots \cdot p_{r+1}^{K_{r+1}} \end{aligned}$$

- **Übung:** Erklären Sie die Verteilung.

■ Modell:

- Ausgangspunkt ist eine Binomialverteilung
- Anzahl der Versuche (N) wird groß
- Wahrscheinlichkeit skaliert mit $p \sim 1/N$

■ Motivation:

- Approximation der Binomialverteilung für viele Versuche
- Zählen von Treffern bei einzeln recht unwahrscheinlichen Ereignissen

Satz

Besitzt X_N eine Binomialverteilung und gilt mit $\lambda > 0$ der Zusammenhang

$$p = \frac{\lambda}{N} \quad \Leftrightarrow \quad \lambda = pN,$$

so gilt:

$$\lim_{N \rightarrow \infty} P(X_N = K) = \frac{\lambda^K}{K!} e^{-\lambda}.$$



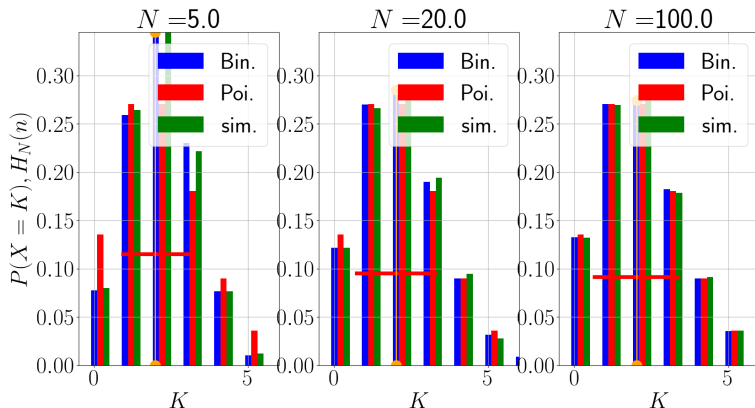
- **Nachweis:** Verwende den Ansatz

$$\begin{aligned}P(X_N = K) &= \frac{N!}{K! (N - K)!} \left(\frac{\lambda}{N} \right)^K \left(1 - \frac{\lambda}{N} \right)^{N-K} \\&= \frac{\lambda^K}{K!} \left(1 - \frac{\lambda}{N} \right)^N \cdot \frac{N(N-1) \cdots (N-K+1)}{N^K \cdot \left(1 - \frac{\lambda}{N} \right)^K} \\&= \frac{\lambda^K}{K!} \left(1 - \frac{\lambda}{N} \right)^N \cdot \frac{1(1 - \frac{1}{N}) \cdots (1 - \frac{K-1}{N})}{\left(1 - \frac{\lambda}{N} \right)^K}\end{aligned}$$

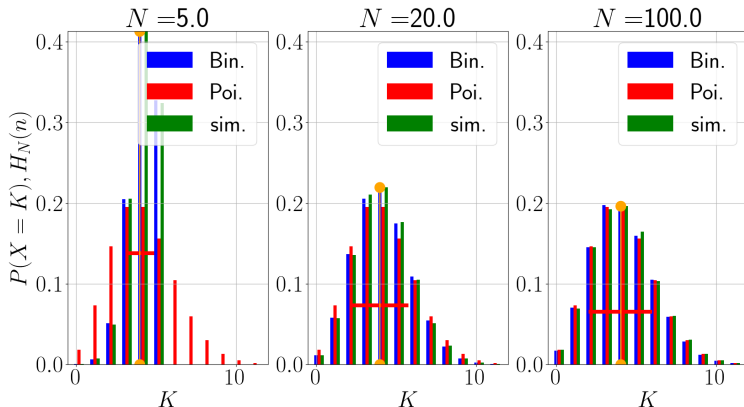
und beachte $\lim_{N \rightarrow \infty}$.



- **Illustration:** Binomial- und Poissonverteilung für $N = 5, 20, 100$ mit $\lambda = 2$;
Simulation von 10 000 Werten



- **Illustration:** Binomial- und Poissonverteilung für $N = 5, 20, 100$ mit $\lambda = 4$;
Simulation von 10 000 Werten



Definition

Bei der *Poissonverteilung* mit Parameter λ nimmt die Zufallsvariable X die Werte $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ mit Wahrscheinlichkeit

$$P(X = K) = \frac{\lambda^K}{K!} e^{-\lambda}, \quad K \in \mathbb{N}_0$$

an.

■ Kenngrößen der Verteilung:³²

$$P(X = K) = \frac{\lambda^K}{K!} e^{-\lambda}$$

$$E(X) = \lambda$$

$$D^2(X) = \lambda$$

$$\varphi(s) = \exp(\lambda(e^{js} - 1))$$

³² **Übung:** Erklären Sie die Ergebnisse anschaulich.



■ Bemerkungen:

- Die Interpretation der Poissonverteilung ergibt sich durch die Grenzwertaussage der Binomialverteilung: Eine Poissonverteilung zählt das Auftreten von Ereignissen, die einzeln relativ unwahrscheinlich sind.
- Die Aussage $E(X) = \lambda$ ist in Verbindung mit $\lambda = pN$ anschaulich nachvollziehbar.

■ Modell:

- Urne mit N Kugeln, von denen M schwarz und $N - M$ weiß sind
- n Kugeln werden ohne Zurücklegen und ohne Beachten der Reihenfolge gezogen
- X beschreibt die Anzahl der gezogenen schwarzen Kugeln

■ Anwendung/Motivation:

- Anzahl der Richtigen beim Lotto
- Auswahl von „guten“ und „schlechten“ Entitäten aus einer Gesamtmenge

■ Beobachtung/„Kuriosität“:³³

- Betrachte das n -fache Ziehen ohne Zurücklegen, wobei $X_i \in \{s, w\}$ das Ergebnis des i -ten Ziehens bezeichnet
- Es folgt:

$$\begin{aligned} P(X_2 = s) &= P(X_2 = s | X_1 = s) P(X_1 = s) \\ &\quad + P(X_2 = s | X_1 = w) P(X_1 = w) \\ &= \frac{M-1}{N-1} \cdot \frac{M}{N} + \frac{M}{N-1} \cdot \frac{N-M}{N} \\ &= \frac{M^2 - M + MN - M^2}{N(N-1)} \\ &= \frac{M}{N} \end{aligned}$$

³³Nach [Kre91, S. 35]

■ Beobachtung/„Kuriosität“:³⁴

- Für das dritte Ziehen ergibt sich:

$$\begin{aligned}P(X_3 = s) &= P(X_3 = s | X_2 = s, X_1 = s)P(X_2 = s, X_1 = s) \\&\quad + P(X_3 = s | X_2 = s, X_1 = w)P(X_2 = s, X_1 = w) \\&\quad + P(X_3 = s | X_2 = w, X_1 = s)P(X_2 = w, X_1 = s) \\&\quad + P(X_3 = s | X_2 = w, X_1 = w)P(X_2 = w, X_1 = w) \\&= \frac{M}{N}\end{aligned}$$

- Es kann gezeigt werden (siehe [Kre91, S. 35]), dass dies auch für weitere Zeitpunkte gilt, da es sich bei der hypergeometrischen Verteilung um eine *austauschbare Verteilung* handelt.

³⁴Nach [Kre91, S. 35]

■ Kenngrößen:

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

$$E(X) = n \frac{M}{N}$$

$$D^2(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$



■ Nachweis³⁵ der Eigenschaften:

- Die erste Aussage folgt aus der Kombinatorik und der Eigenschaft von Laplace-Experimenten.
- Für die zweite Aussage betrachte $X_i = I_{\{s\}}(a_i)$ („ i -te Kugel ist schwarz“). Dann besitzt X_i eine Zweipunktverteilung mit $x_2 = 0$, $x_1 = 1$ und $p = P(X_i = 1) = \frac{M}{N}$ und somit $E(X_i) = p$. Schließlich ist $X = X_1 + \dots + X_n$ und es folgt:

$$E(X) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) = np = n \frac{M}{N}$$

- Die dritte Aussage benötigt etwas mehr Arbeit. Bei Interesse siehe [Kre91]. ■

³⁵Der unter anderem deshalb instruktiv ist, weil er sich im zweiten Punkt eines hilfreichen Tricks aus [Kre91] bedient.

■ Beispiele:

- *Lotto*: Die Wahrscheinlichkeit für 4 Richtige beim „6 aus 49“ (ohne irgendwelche Zusatz-/Superzahlen) beträgt:

$$\frac{\binom{6}{4} \binom{43}{2}}{\binom{49}{6}} \approx 0.000968 \approx 9.7 \cdot 10^{-4}$$

- *Skat*: Die Wahrscheinlichkeit beim Skat alle 4 Buben zu erhalten folgt zu:

$$\frac{\binom{4}{4} \binom{28}{6}}{\binom{32}{10}} \approx 0.00584 \approx 5.8 \cdot 10^{-3}$$

Satz

Besitzt X eine hypergeometrische Verteilung mit Parameter N, M, n , dann gilt mit

$$p = \frac{M}{N}$$

bei festem n :

$$\lim_{N \rightarrow \infty} P(X = K) = \binom{n}{K} p^K (1 - p)^{n-K}$$

Für große N, M entspricht die hypergeometrische Verteilung etwa einer Binomialverteilung mit $p = \frac{M}{N}$ und n -maligem Ziehen.³⁶

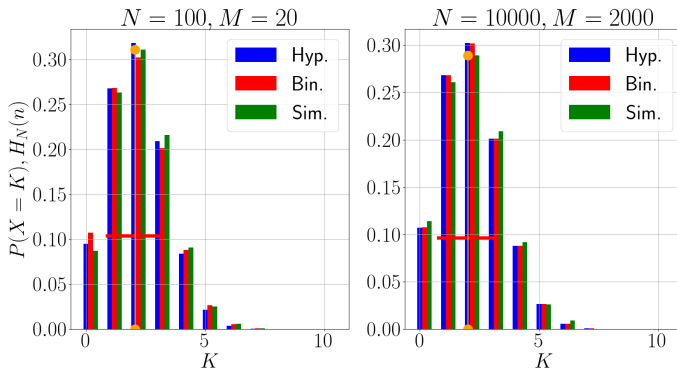
³⁶ **Beachte:** Da $p = M/N$ konstant ist, muss M mit N zunehmen.

- **Nachweis:** Man rechnet:

$$\begin{aligned} P(X = k) &= \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \\ &= \frac{\frac{M \cdots (M-k+1)}{k!} \cdot \frac{(N-M) \cdots (N-M-n+k+1)}{(n-k)!}}{\frac{N \cdots (N-n+1)}{n!}} \\ &= \binom{n}{k} \frac{M \cdots (M-k+1) \cdot (N-M) \cdots (N-M-n+k+1)}{N \cdots (N-n+1)} \end{aligned}$$

Teile Zähler und Nenner durch N^n . $\implies$ Erster Term in Zähler konvergiert gegen p^k , zweiter Term gegen $(1-p)^{n-k}$, Nenner gegen 1. ■

- **Beispiel:** Approximation der hypergeometrischen Verteilung durch eine Binomialverteilung; zudem Simulation mit 1 000 Realisierungen



■ Bemerkungen:

- Die Aussage „folgt nicht einfach“ aus $P(X_i = s) = \frac{M}{N}$, da die Einzelziehungen nicht unabhängig sind.
- Anschauung: Wenn N sehr groß ist (und damit auch $M = pN$ groß), dann fällt das Entnehmen der Kugel nicht sonderlich ins Gewicht; die Wahrscheinlichkeit ist immer noch $\approx \frac{M}{N}$.
- **Übung:** Rechnen Sie nach, dass die Binomialapproximation für das „Lotto 4 Richtige“-Beispiel keine gute Näherung liefert.

4 Diskrete Verteilungen

- Wiederholung
- Zufallsvariablen
- Kenngrößen von Zufallsvariablen
- Wichtige Verteilungen
- **Lernziele**
- Literatur

- Die folgende Aufstellung fasst die zentralen Punkte zusammen.
- Es wird aufgezeigt, welche Punkte nach Bearbeitung des Kapitels klar sein sollten.
- **Hinweise:**
 - Die Auflistung ist nicht vollständig, sondern führt die wichtigsten Aussagen auf; nicht erwähnte Inhalte sind dennoch bedeutsam.
 - Oft enthalten die Nachweise wichtige Ideen; diese also nicht vernachlässigen.
 - Stets versuchen, Gleichungen in Verbindung mit Interpretationen und Anwendungen zu sehen
 - Des weiteren sollten alle kleinen nützlichen Ergänzungen verstanden sein.
 - Es ist immer eine gute Idee, etwas Gelerntes im Rechner umzusetzen. Dies hilft beim Verständnis und schärft das Bewusstsein für mögliche Probleme.



Nach diesem Kapitel sollten als zentrale Punkte klar sein:

- Zufallsvariablen zur „Übertragung“ von Wahrscheinlichkeit
- Motivation der Verwendung von Momenten
- Erwartungswert, Varianz und k -te Momente inkl. Rechenregeln
- Wichtige Verteilungen und ihre Eigenschaften



4 Diskrete Verteilungen

- Wiederholung
- Zufallsvariablen
- Kenngrößen von Zufallsvariablen
- Wichtige Verteilungen
- Lernziele
- Literatur



- [JW02] F. Jondral, A. Wiesler, *Wahrscheinlichkeitsrechnung und stochastische Prozesse*, Teubner, 2002
- [Kre91] U. Krengel, *Wahrscheinlichkeitstheorie und Statistik*, Vieweg, 1991
- [Ren62] A. Rényi, *Wahrscheinlichkeitsrechnung – mit einem Anhang über Informationstheorie*, VEB Verlag der Wissenschaften, 1962
- [Hen13] N. Henze. *Stochastik für Einsteiger – Eine Einführung in die faszinierende Welt des Zufalls*, Springer, 2013
- [Pap91] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd Ed., New York 1991: McGraw-Hill, Inc.
- [Sch15] S. Schäffler, *Mathematik der Information – Theorie und Anwendungen der Shannon-Wiener Information*, Springer, 2015