WAHRSCHEINLICHKEITSTHEORIE UND STATISTIK

Zusammenfassung zur Vorlesung von
Prof. Dr. A.-S. Sznitman

Lukas Cavigelli, Juli 2011
lukasc@ee.ethz.ch

GRUNDBEGRIFFE & ELEMENTARE MODELLE

KOMBINATORIK

Gegeben: zwei endliche Mengen E, F mit $|E| = p$ und $|F| = n$

Variationen mit Wiederholungen:

(Anzahl Abbildungen von E nach F) $= n^p$

- Spezialfall: (Anz. Teilmengen $A \subseteq E$) $= 2^p$

Variationen ohne Wiederholungen: Es sei $n \geq p$.
Anzahl Möglichk. p versch. Objekte auf n Plätze zu verteilen:

$$n \cdot (n-1) \cdot \ldots \cdot (n-p+1) = \frac{n!}{(n-p)!}$$

- Spezialfall: $n = p$: #Permutationen $= n!$

Kombinationen ohne Wiederholungen:
p identische Objekte, n versch. Plätze, $n \geq p$, ein Obj. pro Pos.

$$\#\text{Möglichkeiten} = \binom{n}{p} = \frac{n!}{p! \cdot (n-p)!}$$

Kombinationen mit Wiederholungen:
p identische Obj., n versch. Positionen, mehrere Obj. pro Pos.

$$\#\text{Platzierungen} = \binom{n+p-1}{p}$$

Identitäten des Binomialkoeffizienten:

$$\binom{n}{p} = \binom{n}{n-p}, \sum_{k=0}^{n} \binom{n}{k} = 2^n, \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x+y)^n$$

$$\sum_{k=0}^{n} k \binom{n}{k} = n 2^{n-1}, \binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

$$\binom{n+m}{r} = \sum_{k=0}^{r} \binom{n}{k} \binom{m}{r-k}, n, m \geq r$$

GRUNDRAUM & EREIGNISSE

Grundraum: Die nichtleere Menge Ω, welche alle elementaren Ereignisse enthält, heisst Grundraum.
Elementarereignis: $\omega \in \Omega$ heisst Elementarereignis.
Ereignis: $A \subseteq \Omega$ heisst Ereignis.

WAHRSCHEINLICHKEITSRÄUME

Wahrscheinlichkeit eines Ereignisses:
Jedem Ereignis A wird die Wahrscheinlichkeit $P[A] \in [0,1]$ zugeordnet. Dabei gilt $P[\Omega] = 1$.
Ist $A = \bigcup_{i=1}^{k} A_i$ und A_i paarweise disjunkt, so gilt:

$$P[A] = \sum_{i=1}^{k} P[A_i]$$

Wahrscheinlichkeitsraum:
Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathcal{F}, P)$, wobei Ω der Grundraum, $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ eine Kollektion von Ereignissen und P eine Funktion $A \in \mathcal{F} \to P[A] \in [0,1]$ ist. Dazu setzen wir voraus, dass $\mathcal{F}$ und P folgendes erfüllen:
$\mathcal{F}$ ist eine ***σ*-Algebra**, d.h. es gilt:

- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$
- Für jede Folge $A_n, n \geq 1$ mit $A_n \in \mathcal{F}$ gilt $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

P ist eine Wahrscheinlichkeit, d.h. es gilt:

- $P[\Omega] = 1$
- Für A_n, $n \geq 1$ eine Folge von Ereignissen, welche paarweise disjunkt sind ($A_i \cap A_j = \emptyset \ \forall i \neq j$) gilt ($\sigma$-Additivität):
$$P[\cup_{n=1}^{\infty} A_n] = \sum_{n=1}^{\infty} P[A_n]$$

Folgerungen aus der Definition:

- $P[\emptyset] = 0$
- $A \in \mathcal{F} \Rightarrow P[A^C] = 1 - P[A]$
- $A_1, \ldots, A_n \in \mathcal{F}$ endl. Folge mit $A_i \cap A_j = \emptyset \ (1 \leq i \neq j \leq n)$, dann gilt: $P[\cup_{i=1}^{n} A_i] = \sum_{i=1}^{n} P[A_i]$
- $A \subseteq B, \ A, B \in \mathcal{F} \Longrightarrow P[A] \leq P[B]$
- $A \subseteq B \Longrightarrow P[B \setminus A] = P[B] - P[A]$
- Seien $A, B, C \in \mathcal{F}$, dann gilt (Prinzip der ***Inklusion/Exklusion***):
$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$
$$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B] - P[B \cap C] - P[C \cap A] + P[A \cap B \cap C]$$
- ***De-Morgan***
$$A \cap B = (A^C \cup B^C)^C, \qquad A \cup B = (A^C \cap B^C)^C$$
- ***Kettenregeln***
$$P[A \cap B] = P[A|B] \cdot P[B] = P[B|A] \cdot P[A]$$
$$P[A \cap B \cap C] = P[C|A \cap B] \cdot P[A \cap B] = P[C|A \cap B] \cdot P[B|A] \cdot P[A]$$
$$P[A] = P[A \cap B] + P[A \cap B^C]$$

LAPLACE-MODELL

Laplace-Modell:
Ist der Grundraum Ω *endlich*, $\mathcal{F} = \mathcal{P}[\Omega]$ und haben alle *Elementarereignisse $\omega \in \Omega$ dieselbe Wahrscheinlichkeit*

$$p(\omega) = P[\{\omega\}] = 1/|\Omega|$$

so heisst der W'keitsraum $(\Omega, \mathcal{F}, P)$ Laplace-Modell.
Ist $A \subseteq \Omega$ (d.h. $A \in \mathcal{F}$), so gilt:

$$P[A] = \frac{|A|}{|\Omega|} = \frac{\#\text{ günstige}}{\#\text{ mögliche}}$$

BEDINGTE WAHRSCHEINLICHKEIT

Bedingte Wahrscheinlichkeit:
Sei $(\Omega, \mathcal{F}, P)$ ein W'keitsraum und $A, B \in \mathcal{F}$ Ereignisse mit $P[B] > 0$, dann heisst:

$$P[A|B] = P[A \cap B]/P[B]$$

die *bedingte Wahrscheinlichkeit* von A gegeben B.
Folgerungen aus der Definition:

- $P[A^C|B] = 1 - P[A|B]$ weil $P[\cdot|B]$ ein W'keitsmass.
- i.A. $P[A|B] \neq 1 - P[A|B^C]$
- wenn A_i paarweise disjunkt: $P[(\cup_i A_i)|B] = \sum_i P[A_i|B]$
- $P[(\cap_i A_i)|B] \xleftrightarrow{unabh.} \prod_i P[A_i|B]$

Formel der totalen Wahrscheinlichkeit:
Sei $B_1, \ldots, B_n \in \mathcal{F}$ eine Zerlegung von Ω, d.h. $\Omega = B_1 \cup \ldots \cup B_n$ mit $P[B_i] > 0 \ \forall i \in \{1, \ldots, n\}$. Dann gilt für ein Ereignis $A \in \mathcal{F}$:

$$P[A] = \sum_{i=1}^{n} P[A|B_i] P[B_i]$$

Für $A \in \mathcal{F}$ und f_Y stetig gilt analog:

$$P[A] = \int_{E_Y} P[A|Y = y] f_Y(y) \, dy$$

Formel von Bayes: Sei $B_1, \ldots, B_n \in \mathcal{F}$ eine Zerlegung von Ω mit $P[B_i] > 0 \ \forall i \in \{1, \ldots, n\}$ und $A \in \mathcal{F}$ ein Ereignis mit $P[A] > 0$.

$$P[B_i|A] = \frac{P[A|B_i] \cdot P[B_i]}{P[A]} = \frac{P[A|B_i] \cdot P[B_i]}{\sum_{j=1}^{n} P[A|B_j] \cdot P[B_j]}$$

$P[B_i]$ heisst die *a priori W'keit* der Ursache B_i.
$P[B_i|A]$ heisst die *a posteriori W'keit* der Ursache B_i (gegeb. A)

UNABHÄNGIGKEIT

Unabhängigkeit zweier Ereignisse:
Sei $(\Omega, \mathcal{F}, P)$ ein W'keitsraum. Seien zwei Ereignisse $A, B \in \mathcal{F}$ (mit möglicherweise Null W'keit) heissen *unabhängig*, wenn:

$$P[A \cap B] = P[A] \cdot P[B]$$

Ist $0 < P[A] < 1$, so sind A und A^C *nie unabhängig*.
Sind A und B *unabhängig*, so gilt

$$P[A|B] = P[A]$$

Unabhängigkeit zweier Zufallsvariablen:
Sei $(\Omega, \mathcal{F}, P)$ ein W'keitsraum. Seien X, Y zwei Zufallsvar. auf Ω. X und Y heissen *unabhängig*, wenn $\forall a, b \in \mathbb{R}$ gilt:

$$P[\{X \leq a\} \cap \{Y \leq B\}] = P[X \leq a] \cdot P[Y \leq b]$$

Unabhängigkeit einer Kollektion von Ereignissen:
Sei $A_1, \ldots, A_n$ eine Kollektion von Ereignissen (d.h. $A_i \in \mathcal{F} \ \forall i$)
Die Kollektion heisst unabh., falls

$$\forall 1 \leq i_1 < i_2 < \cdots < i_k \leq n \text{ mit } 2 \leq k \leq n \text{ gilt:}$$
$$P[A_{i_1} \cap \ldots \cap A_{i_k}] = P[A_{i_1}] \cdot \ldots \cdot P[A_{i_k}]$$

ZUFALLSVARIABLEN, VERTEILUNGEN, DICHTE

ZUFALLSVARIABLE

Zufallsvariable:
Sei $(\Omega, \mathcal{F}, P)$ ein W'keitsraum. Eine reelle *Zufallsvariable* ist eine Abb. $X: \Omega \to \mathbb{R}$, so dass für $a \in \mathbb{R}$ gilt:

$$\{\omega \in \Omega, X(\omega) \leq a\} \in \mathcal{F}$$

Notation: Man schreibt $\{X \leq a\}$ für $\{\omega \in \Omega, X(\omega) \leq a\}$.

VERTEILUNG EINER ZUFALLSVARIABLEN

Verteilungsfunktion:
Sei X eine Zufallsvariable auf $(\Omega, \mathcal{F}, P)$. Die Verteilungsfunktion von X ist die Funktion $F_X: \mathbb{R} \to [0,1]$ definiert durch:

$$F_X(a) = P[\{X \geq a\}], \qquad a \in \mathbb{R}$$

Eigenschaften von Verteilungsfunktionen:
Für die Verteilungsfunktion F_X der Zufallsvariablen X gilt:

- F_X monoton wachsend auf $\mathbb{R}$ mit Werten in $[0,1]$
- $\lim_{a \to +\infty} F_X(a) = 1$ und $\lim_{a \to -\infty} F_X(a) = 0$
- F_X ist rechtsseitig, d.h. $F_X(a) = \lim_{h \to 0^+} F_X(a+h)$
- Für $a \in \mathbb{R}$ gilt: $P[\{X < a\}] = F_X(a^-) = \lim_{h \to 0^+} F_X(a-h)$
- Für $a \in \mathbb{R}$ gilt: $P[\{X = a\}] = F_X(a) - F_X(a^-)$
- $P[\{X > a\}] = 1 - F_X(a)$ und $P[\{X \geq a\}] = 1 - F_X(a^-)$
- $P\big[X \in [a,b]\big] = P[a \leq X \leq b] = F_X(b) - F_X(a^-)$

Diskrete Verteilung:
Die Verteilungsfunktion einer diskreten Zufallsvariablen X lässt sich mittels der W'keitsfunktion $P[X = x_i] = p_i \ \forall i$ bestimmen:

$$F_X(x) = P[X \leq x] = \sum_{x_i \leq x} P[X = x_i] = \sum_{x_i \leq x} p_i$$

Dabei gelten folgende Eigenschaften:

- $\sum_{i=1}^{\infty} P[X = x_i] = 1$
- Für $a, b \in \mathbb{R}, a \leq b$ gilt: $P[\{a < X \leq b\}] = F_X(b) - F_X(a)$

Stetige Verteilung:
$X: \Omega \to \mathbb{R}$ Zufallsvariable auf $(\Omega, \mathcal{F}, P)$ hat eine stetige Verteilung, wenn $f_X(x) \geq 0$ und $\int_{-\infty}^{\infty} f_X(x) dx = 1$.
Existiert eine nicht-negative Funktion $f_X(x)$ auf $\mathbb{R}$, so dass:

$$F_X(x) = P[X \leq x] = \int_{-\infty}^{x} f_X(u) \, du, \qquad x \in \mathbb{R}$$

so heisst f_X die Dichtefunkt. von X. Es muss gelten:

$$f_X(x) = \frac{d}{dx} F_X(x), \quad \int_{-\infty}^{\infty} f_X(x) dx = 1$$

Besitzt X eine Dichtefunktion f_X ($\Leftrightarrow F_X$ stetig und stückw. diff'bar), so gilt:

$$\int_a^b f_X(x) dx = P[a < X < b] = P[a \leq X \leq b] = P[a < X \leq b] = P[a \leq X < b]$$

TRANSFORMATION EINER STETIGEN ZV

Sei $g: \mathbb{R} \to \mathbb{R}$ eine Zufallsvariable (ZV) auf $(\Omega, \mathcal{F}, P)$. Es sei $Y = g(X)$. Ist g genügend schön, so ist auch Y eine ZV.
Ist g streng monoton steigend und stetig, so existiert g^{-1} und wir erhalten die Verteilungsfunktion von Y:

$$F_Y(a) = P[Y \leq a] = P[g(X) \leq a] = P[X \leq g^{-1}(a)] = F_X(g^{-1}(a))$$

Für die Dichte erhalten wir unter der Annahme, dass g diff'bar und X eine Dichte besitzt:

$$f_Y(a) = f_X(g^{-1}(a)) \cdot \frac{1}{|g'(g^{-1}(a))|}$$

ZENTRALER GRENZWERTSATZ

Sind X_i *independent and identically distributed (iid)* Zufallsvariablen, und $S = X_1 + \cdots + X_n$ mit $E[S] = n \cdot E[X_i]$ und $\mathrm{Var}[S] = n \cdot \mathrm{Var}[X_i]$, so gilt:

$$P\left[\frac{S - E[S]}{\sqrt{\mathrm{var}[S]}} \leq z\right] \xrightarrow{n \to \infty} \mathcal{N}(0,1)$$

ERWARTUNGSWERT, VARIANZ, ...

FÜR DISKRETE ZUFALLSVARIABLEN

Erwartungs- bzw. Mittelwert:

$$\mu = E[X] = \sum_i x_i P[X = x_i] = \sum_{x \in E_X} x p_X(x)$$

Varianz $\sigma^2 = \mathrm{Var}[X]$ und **Standardabweichung** $\sigma = \sqrt{\mathrm{Var}[X]}$:

$$\sigma^2 = \mathrm{var}[X] = \sum_i (x_i - \mu)^2 P[X = x_i] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

FÜR STETIGE ZUFALLSVARIABLEN

Erwartungs- bzw. Mittelwert:

$$\mu = E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx$$

Varianz $\sigma^2 = \text{var}[X]$ und **Standardabweichung** $\sigma = \sqrt{\text{var}[X]}$:

$$\sigma^2 = \text{var}[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx = E[(X-E[X])^2] = E[X^2] - E[X]^2$$

EIGENSCHAFTEN

Seien die Zufallsvariable X und Y *unabhängig*, so gilt:

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Symmetrie der Varianz:

$$\text{var}[X] = \text{var}[-X]$$

Satz des totalen Erwartungswertes:

Sei $\Omega = A_1 \cup A_2 \cup \ldots \cup A_n$ mit $A_i \cap A_j = \emptyset \forall i \neq j$ eine Zerlegung des Grundraumes, sowie X eine Zufallsvar. Dann:

$$E[X] = \sum_{k=1}^{n} P[A_k] \cdot E[X|A_k]$$

ERWARTUNGSWERT EINER FUNKTION

Sei X eine ZV und $Z = g(X)$. Der Erwartungswert $E[Z]$ ist

- wenn X eine *diskrete* ZV:

$$E[Z] = E[g(X)] = \sum_i g(x_i) \cdot P[X = x_i]$$

- wenn X eine *stetig* ZV:

$$E[Z] = E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\,dx$$

Für zwei *unabhängige* Zufallsvariablen X und Y und zwei Funktionen f ung g gilt:

$$E[f(X) \cdot g(Y)] = E[f(X)] \cdot E[g(Y)]$$

ERWARTUNGSWERT & VARIANZ EINER LIN. F.

Sei $Z = a \cdot X + b$, so gilt:

$$E[Z] = E[a \cdot X + b] = a \cdot E[X] + b$$
$$\sigma^2 = \text{var}[Z] = \text{var}[a \cdot X + b] = a^2\,\text{var}[X]$$

ZWEIDIMENSIONALE VERTEILUNGEN

Die Verteilung von zwei Zufallsvariablen X und Y:

$$F(x,y) = P[X \leq x, Y \leq y]$$

Dabei gelten folgende *Eigenschaften*:

- $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$
- $\lim_{x\to\infty,\ y\to\infty} F(x,y) = 1$

STETIGE ZWEIDIMENSIONALE VERTEILUNG

Gemeinsame Dichte:

X, Y ZV: $\Omega \to \mathbb{R}$ auf $(\Omega, \mathcal{F}, P)$ und besitzen eine gemeinsame stetige Verteilung, falls $f_{X,Y}(x,y) \geq 0$ auf $\mathbb{R}^2$ existiert, wobei $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx\,dy = 1$, so dass gilt:

$$P\big[X \in [a_1, b_1], Y \in [a_2, b_2]\big] = \int_{a_2}^{b_2}\int_{a_1}^{b_1} f_{X,Y}(x,y)dx\,dy$$

wobei $-\infty \leq a_1 \leq b_1 \leq \infty$ und $-\infty \leq a_2 \leq b_2 \leq \infty$. Dann heisst $f_{X,Y}(x,y)$ gemeinsame Dichte von X und Y.

Verteilungsfunktion:

$$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(\tilde{x},\tilde{y})d\tilde{x}\,d\tilde{y}$$

Randdichten der gemeinsamen Dichte:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

DISKRETE ZWEIDIM. VERTEILUNG

W'keitsfunktion: (normiert!)

$$P_{X,Y}(x,y) = \begin{cases} p_{x,y}, & X = x, Y = y \\ 0, & sonst \end{cases}$$

Verteilungsfunktion:

$$F_{X,Y}(x,y) = \sum_{x \leq X}\sum_{y \leq Y} P_{X,Y}(x,y)$$

W'keitsfunktionen der einzelnen ZV:

$$P_X(x) = \sum_y P_{X,Y}(x,y), \qquad P_Y(y) = \sum_x P_{X,Y}(x,y)$$

Randverteilungen:

$$F_X(x) = \sum_{x \leq X} P_X(x) = \sum_{x \leq X}\sum_{y} P_{X,Y}(x,y)$$

UNABHÄNGIGKEIT

Seien X, Y zwei ZV und g, h zwei Funktionen, dann:

X, Y unabhängig
$$\Longrightarrow F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$$
$$\Longrightarrow P_{X,Y}(x,y) = P_X(x) \cdot P_Y(y)$$
$$\Longrightarrow f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$
$$\Longrightarrow E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)]$$

ERWARTUNGSWERT

$$E[X \cdot Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f_{X,Y}(x,y)dx\,dy$$

Sei $Z = g(X,Y)$ eine ZV, so gilt für deren Erwartungswert:

$$E[Z] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dx\,dy$$

KOVARIANZ

Die **Kovarianz** beschreibt den Grad der Abhängigkeit zw. 2 ZV.

$$\text{cov}(X,Y) \coloneqq E[(X - E[X])(Y - E[Y])] = \sigma_{X,Y} = E[X \cdot Y] - E[X] \cdot E[Y]$$

Korrelationsfaktor:

$$\text{cor}(X,Y) = \rho = \frac{\sigma_{X,Y}}{\sigma_X \cdot \sigma_Y} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}} \in [-1,1]$$

Eigenschaften der Kovarianz:

- X, Y unabhängig $\Longrightarrow \text{cov}(X,Y) = 0$
- $\text{cov}(X,Y) = 0 \Longrightarrow X, Y$ unkorreliert (nicht zwingend unabh.)
- $\text{cov}(X,X) = \text{var}(X)$
- $\text{cov}(aX + b, Y) = a\,\text{cov}(X,Y)$
- $\text{cov}(X + Y, Z) = \text{cov}(X,Z) + \text{cov}(Y,Z)$

BESTE LINEARE PROGNOSE

Die **beste lineare Prognose** von Y durch X im Sinn der Minimierung des erwarteten Fehlers $E[(Y-Z)^2]$ unter allen linearen Transformationen $\alpha X + \beta$ ist:

$$Y \approx Z = \frac{\text{cov}[X,Y]}{\text{var}[X]}(X - E[X]) + E[Y]$$

Minimierter Fehler: $E[(Y-Z)^2] = \text{var}[Y]\,(1 - \text{cor}[X,Y]^2)$

SPEZIELLE VERTEILUNGEN

BERNOULLI-VERTEILUNG $X \sim be(p)$

Die BenoullIverteilung ist die Verteilung von einem zufälligen 0-1-Experiment mit Erfolgsparameter $p \in [0,1]$ und $X: \Omega \to \{0,1\}$.

$$P[X = 1] = p, \qquad P[X = 0] = 1 - p$$

Verteilungsfunktion: $F(k) = P[X \leq k] = \begin{cases} 0, & k < 0 \\ 1 - p, & 0 \leq k < 1 \\ 1, & 1 \leq k \end{cases}$

Erwartungswert: $E[X] = p$

Varianz: $\text{var}[X] = p(1-p)$

BINOMIALVERTEILUNG $X \sim bn(n,p)$

Die Verteilung der totalen Anzahl von Erfolgen von n unabhängigen 0-1-Experimenten heisst Binomialverteilung der Länge n mit Erfolgparameter p.

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0,1,\ldots,n\}$$

$F(k) = P[X \leq k] = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$

$E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = nE[X_1] = np$

$\text{var}[X] = \text{var}[X_1 + \cdots + X_n] = \cdots = n\,\text{var}[X_1] = np(1-p)$

Stabilität der Binomialverteilung:

Sei $X \sim \text{bn}(n_1, p)$ und $Y \sim \text{bn}(n_2, p)$ unabhängig, dann:

$$X + Y \sim \text{bn}(n_1 + n_2, p)$$

GEOMETRISCHE VERTEILUNG $X \sim geom(p)$

Anzahl Versuche bis zum ersten Erfolg in einer unendl. Folge von unabhängigen 0-1-Experimenten mit Erfolgsparam. p.

$$P[X = k] = p(1-p)^{k-1}, \qquad k \geq 1$$

$F(k) = P[X \leq k] = p\sum_{i=1}^{k}(1-p)^{i-1} = 1 - (1-p)^k$

$E[X] = \cdots = -p\frac{d}{dp}\left(\sum_{k=0}^{\infty}(1-p)^k\right) = -p\frac{d}{dp}\left(\frac{1}{p}\right) = \frac{1}{p}$

$\text{var}(X) = \frac{1}{p^2} - \frac{1}{p}$

Gedächtnislosigkeit: $P[T = n_0 + n | T > n_0] = P[T = n] \forall n, n_0$

HYPERGEOMETRISCHE VERTEILUNG

W'keit dafür, dass bei N gegebenen Elementen, von denen M die gewünschte Eigenschaft besitzen, beim Herausgreifen von n Probestücken genau k Treffer erzielt werden, d.h. die W'keit für $X = k$ Erfolge in n Versuchen.

$$P[X = k] = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$F(k) = P[X \leq k] = \sum_{i=0}^{k} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$

$E[X] = n\frac{M}{N}$, $\text{var}(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$

POISSON-VERTEILUNG $X \sim pois(\lambda)$

Approximation für Binomialverteilung für

$$n \to \infty, \quad p \to 0, \quad np = \lambda \in (0,\infty)$$

$P[X = k] = e^{-\lambda}\frac{\lambda^k}{k!}, \ k \geq 0$

$F(k) = P[X \leq k] = e^{-\lambda}\sum_{i=0}^{k}\frac{\lambda^i}{i!}$, $E[X] = \lambda$, $\text{var}(X) = \lambda$

Stabilität: $X \sim \text{pois}(\lambda)$, $Y \sim \text{pois}(\mu)$, dann: $X + Y \sim \text{pois}(\lambda + \mu)$

GLEICHVERTEILUNG $X \sim \mathcal{U}(a,b)$

auf dem Intervall $[a,b]$, $a < b$

Dichte: $f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}$

Verteilung: $F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$

$E[X] = \frac{a+b}{2}$, $\text{var}[X] = \frac{1}{12}(b-a)^2$

EXPONENTIALVERTEILUNG $X \sim exp(\lambda)$

Stetiges Analogon zur geometrischen Verteilung.

Dichte: $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

Verteilung: $F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

$E[X] = \frac{1}{\lambda}$, $\text{var}[X] = \frac{1}{\lambda^2}$

Gedächtnislosigkeit: $P[X \geq x_0 + x | X \geq x_0] = P[X \geq x]$

GAUSS'SCHE NORMALVERTEILUNG $X \sim \mathcal{N}(\mu, \sigma^2)$

μ: Zentrierungsparameter, σ: Breiteparameter

Dichte: $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

Verteilung: $F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx$

$E[X] = \mu$, $\quad \text{var}[X] = \sigma^2$

Wendepunkte der Dichte: $x = \mu \pm \sigma$

Maximum: $f(x_{max} = \mu) = \frac{1}{\sqrt{2\pi}\sigma}$

Summe unabhängiger Normalverteilungen:

Seien alle $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ unabhängig und $Y = \sum_i \alpha_i X_i$, dann:

$$Y \sim \mathcal{N}, \qquad E[Y] = \sum_i \alpha_i E[X_i], \qquad \text{var}[Y] = \sum_i \alpha_i^2\,\text{Var}[X_i]$$

STANDARDNORMALVERTEILUNG $X \sim \mathcal{N}(0,1)$

Gauss'sche Normalverteilung mit $\mu = 0, \sigma = 1$.

Dichte: $\varphi(x) = f(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$

Verteilung: $\Phi(x) = F(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{x^2}{2}\right)dx$

Eigenschaften von φ und Φ:

$$\varphi(x) = \varphi(-x), \qquad \Phi(-x) = 1 - \Phi(x)$$

Umrechnung:

Eine normalverteilte Zufallsvariable X mit Parametern μ und σ lässt sich mit Variablentrafo in die Standnormalvert. überführen

$$U = \frac{X - \mu}{\sigma}, \qquad X \sim \mathcal{N}(\mu, \sigma^2), \qquad U \sim \mathcal{N}(0,1)$$

STATISTIK

MAXIMUM-LIKELIHOOD-METHODE

Kurzfassung: mit zu best. Param. L und Messwerten $x_1, \dots, x_n$

$$\hat{L} = \arg\max_{L} f(x_1, \dots, x_n, L) = \arg\max_{L} \prod_{i=1}^{n} f(x_i, L)$$

$$\Leftrightarrow \hat{L} = \arg\max_{L} \ln\big(f(x_1, \dots, x_k, L)\big) = \arg\max_{L} \sum_{i=1}^{n} \ln\big(f(x_i, L)\big)$$

$$\Leftrightarrow \frac{\partial}{\partial \hat{L}}\left(\prod_{i=1}^{n} f(x_i, \hat{L})\right) = 0 \Leftrightarrow \frac{\partial}{\partial \hat{L}}\left(\sum_{i=1}^{n} \ln\left(f(x_i, \hat{L})\right)\right) = 0$$

Ausführlich: Bei der Maximum-Likelihood-Methode wird von einer Zufallsvariablen X aus gegangen, deren Dichte- bzw. W'keitsfunktion f von einem Parameter q abhängt. Liegt eine einfache Zufallsstichprobe mit n unabhängigen und identisch verteilten Realisationen vor, so lässt sich die Dichtefunktion btw. W'keitsfunktion wie folgt faktorisieren („gemeinsame Dichte"):

$$f(x_1, x_2, \dots, x_n; q) = \prod_{i=1}^{n} f_{X_i}(x_i; q)$$

Statt nun für einen festen Parameter q die Dichte für beliebige Werte $x_1, \dots, x_n$ auszuwerten, kann umgekehrt für beobachtete und somit feste Realisationen $x_1, \dots, x_n$ die Dichte als Funktion von q betrachtet werden. Dies führt im Falle der Zufallsvariable X zur Likelihood-Funktion:

$$L(q) = \underbrace{\prod_{i=1}^{n} f_{X_i}(x_i; q)}_{stetig} = \underbrace{\prod_{i=1}^{n} P_{X_i}(x_i; q)}_{diskret}$$

Wird diese Funktion in Abhängigkeit von q maximiert, so erhält man die Maximum-Likelihood-Schätzung für q. Es wird also der Wert von q gesucht, bei dem die Stichprobenwerte $x_1, \dots, x_n$ die grösste Dichtefunktion haben. Der Maximum-Likelihood-Schätzer ist in diesem Sinne der plausibelste Parameterwert für die Realisierungen $x_1, \dots, x_n$ der Zufallsvariablen X. Die Maximierung dieser Funktion erfolgt, indem man die erste Ableitung nach q bildet und diese dann Null setzt. Da dies bei Dichtefunktionen mit komplizierten Exponentenausdrücken sehr aufwändig werden kann, wird häufig die logarithmierte Likelihood-Funktion verwendet. Die hat an denselben Stellen ihr Maximum und ist einfacher zu berechnen.

$$\ell(q) = \ln\left(\prod_{i=1}^{n} f_{X_i}(x_i; q)\right) = \sum_{i=1}^{n} \ln\left(f_{X_i}(x_i; q)\right)$$

bzw. $\ell(q) = \ln\left(\prod_{i=1}^{n} P_{X_i}(x_i; q)\right) = \sum_{i=1}^{n} \ln\left(P_{X_i}(x_i; q)\right)$

$$\frac{d}{dq}\ell(q) = 0 \to q = \cdots$$

CHI-QUADRAT-TEST (χ^2)

Dieser Test wird verwendet, um zu prüfen, ob die Resultate $N_i \ (1 \le i \le k)$ eines Experimentes mit

$$N = \textstyle\sum_{i=1}^{k} N_i \text{ Durchläufen}$$

einer diskreten Verteilung $F_0(x)$ (Nullhyptohese) entsprechen. Dazu wird zuerst der Prüfwert $\hat{\chi}^2$ mittels der theoretischen Wahrscheinlichkeit p_i $(1 \le i \le k)$ aus $F_0(x)$ berechnet.

$$\hat{\chi}^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i}$$

Dann wählen wir die Signifikanzzahl (Irrtumswahrscheinlichkeit) α (z.B. 0.05) und bestimmen die kritische Grenze C durch

$$P[\chi^2 \le C] = 1 - \alpha$$

(mittels Tabelle).

Die Anzahl **Freiheitsgrade** ist dabei $f = k -$„Anzahl Parameter, die mit denselben Daten geschätzt wurden".

Die Nullhyptohese H_0 verwerfen wir, falls $\hat{\chi}^2 > C$ ist.

Für $\hat{\chi}^2 \le C$ behalten wir die Nullhypothese bei.

Fehler erster Art:

Nullhypothese wird abgelehnt, obwohl sie richtig ist. Die W'keit für diesen Fehler ist

$$\alpha = \int_C^\infty f_{f,0}(x)\, dx$$

$f_{f,0}$ beschreibt die Chi-Dichte der Nullhypothese.

Fehler zweiter Art:

Nullhypothese wird akzeptiert, obwohl sie falsch ist. Je grösser die Macht, desto kleiner die W'keit des Fehlers zweiter Art.

$$Macht = \int_C^\infty f_{f,1}(x)\, dx = 1 - F_{f,1}(C)$$

$f_{f,1}$ beschreibt dabei die Chi-Dichte der Alternativhypothese.

ANDERES

Satz von DeMoivre-Laplace (Serie 11):

$$S_n \sim Bn(n, p), \qquad 0 < p < 1$$

$$\lim_{n\to\infty} P\left[\frac{S_n - np}{\sqrt{np(1-p)}} < z\right] = \Phi(z)$$

Cauchy-Verteilung:

$$f(x) = \frac{1}{\pi}\frac{s}{s^2 + (x-t)^2}, \qquad s > 0, \qquad -\infty < t < \infty$$

$$F(x) = P[X < x] = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-t}{s}\right)$$

Roboter-Problem: (S5A4, A3)

balbal

Integrale:

$$\int_0^\infty t^n e^{-t} dt = n!$$

$$\int_0^\infty t e^{-c\,t^2} dt = \frac{1}{2c}, \qquad c > 0$$

Inkulsion/Exklusion S2

$$P\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \cdots \le i_k \le n} P[A_{i_1} \cap \dots \cap A_{i_k}]$$

Allgemein: Bei Problemen mit dem Komplement erneut versuchen!

Geom. Summe / Reihe:

$$\sum_{k=0}^{n} q^k = \frac{1 - q^{n+1}}{1 - q}, \qquad \xrightarrow{n\to\infty,\ |q|<1} \frac{1}{1-q}$$

$$P[U \cdot V < x] = \int_0^1 \int_{\substack{v \in [0,1] \\ v \le x/u}} f_V(v)\, dv\, f_U(u) du$$

Integralsatz

Wahrscheinlichkeitstheorie

UND STATISTIK

Zusammenfassung zur Vorlesung von
Prof. Dr. A.-S. Sznitman

Lukas Cavigelli, Juli 2011
lukasc@ee.ethz.ch

WAHRSCHEINLICHKEIT

Beispiel 1: Wurf einer Nadel auf einen Parkettboden.
Man wirft mehrmals eine Nadel der Länge L. Die Lattenbreite des Bodens ist $2L$. Das Experiment hat wichtige Eigenschaften:
- Zufall: manchmal ein Latte getroffen, manchmal zwei
- Stabile relative Häufigkeit des Treffens von 1er oder 2 Latten

n Wiederholungen des wurfes, n_1: Anzahl der Treffen der 1. Latte, n_2: Anzahl der Treffen der 2. Latte
Relative Häufigkeit: $\frac{n_2}{n} \approx$ feste Zahl für grosse $n \approx \frac{1}{\pi}$
Beispiel 2: Kodierung von Informationen. Sender sendet ein 0/1-Bit pro Zeiteinheit. Kanal kann nur 0.8 Bit pro Zeiteinh. übertragen. Idee: Kenntnis der stat. Eigenschaften des Senders → günstige Kodierung der Nachricht.
Idee (Huffmann-Codes): Gewisse Blöcke von Bits erscheinen selten. solche Blöcker werden durch viele Bits kodiert. Blöcke die oft auftreten werden durch wenige Bits kodiert. Ohne Informationsverlust.

GRUNDBEGRIFFE & ELEMENTARE MODELLE

GRUNDRAUM & EREIGNISSE

Grundraum: $\Omega \neq \emptyset$. Kollektion aller elementarer Ergebnisse des Experiments. Ω wird so gewählt, dass $\omega \in \Omega$ die maximale Information des Experiments enthält.
Elementarereignis: $\omega \in \Omega$

Bsp. 1: Wurf eines Würfels
$\Omega = \{1,2,3,4,5,6\}$.

Bsp. 2: Wurf einer Nadel:
$a =$ Abstand der Mitte der Nadel zur nächsten Latte.
$\theta =$ Winkel der Nadel zur Richtung der Latten.
$\Omega = [0, L] \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right] = \left\{(a, \theta): 0 \leq a \leq L, -\frac{\pi}{2} < \theta \leq \frac{\pi}{2}\right\}$

Ereignis: Teilmenge $A \subseteq \Omega$ (Grundraum)
Ereignis $A \xleftrightarrow{modelliert}$ Liegt die Realisation $\omega \in \Omega$ des Exp. in A?
Beispiel 1: Wurf eines Würfels. Frage: Zeigt der Würfel eine gerade Zahl? Ereignis: $A = \{2,4,6\} \subseteq \Omega = \{1,2,3,4,5,6\}$
Frage: Zeigt der Würfel ein Vielfaches von 3? Ereig.: $A = \{3,6\}$
Beispiel 2: Wurf einer Nadel auf Parkett.
Frage: Trifft die Nadel zwei Latten?
Ereignis: $A = \left\{(a, \theta) \in [0, L] \times \left(\frac{-\pi}{2}, \frac{\pi}{2}\right], \frac{L}{2}|\sin(\theta)| > a\right\}$

Bemerkung: Die mathem. Best. der Modells ist sehr wichtig!

Beispiel (Bertrand-Paradoxon): Ausgangslage: Einheitskreis mit Gleichseitigem Dreieck. Frage: Wie gross ist die W'keit bei der Wahl einer zufälligen Sehne, eine mit Länge $> \sqrt{3}$ zu wählen? Mögliche Antworten: $\frac{3}{4}, \frac{1}{3}, \frac{1}{2}$

WAHRSCHEINLICHKEITSMASS & -RÄUME

Zufallsexperiment → stabile relative Häufigkeiten
n: Anzahl Wiederholungen des Experiments
n_A: Anzahl Autreten des Ereignisses A
Für grosse n wird $\frac{n_A}{n}$ eine feste Zahl in $[0,1]$
Definition W'keitsmass:
Jedem Ereignisordnen wir eine Zahl zw. 0 und 1 zu:
$$A \to P(A) \in [0,1]$$
Nach Definition: $n_\Omega = n$ (Ω tritt immer auf)
$$P(\Omega) = 1$$
Falls $A = A_1 + A_2 + \cdots + A_k$ (d.h. $A = \bigcup_{i=1}^k A_i$) gilt:
$$\frac{n_A}{n} = \frac{n_{A_1}}{n} + \frac{n_{A_2}}{n} + \cdots + \frac{n_{A_k}}{n}$$
$P(A) = P(A_1) + P(A_2) + \cdots + P(A_k)$ für $A = A_1 + \cdots + A_k$
Definition Wahrscheinlichkeitsraum & σ-Algebra:
Ein Wharscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$, wobei:
- Ω: Grundraum, $\Omega \neq \emptyset$
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$: Eine Kollektion von Ereig. der Potenzmenge von Ω
- P: W'keitsmass, Funktion $A \in F \to P(A) \in [0,1]$

Voraussetzung für $\mathcal{F}$: $\mathcal{F}$ ist eine σ-Algebra, also
- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- Für jede Folge $A_n, n \geq 1, A_n \in \mathcal{F}$ gilt:

$$\bigcup_{i=1}^\infty A_n (\stackrel{\text{def}}{=} \{\omega \in \Omega, \exists i \geq 1, \omega \in A_i\}) \in \mathcal{F}$$
Voraussetzung für P:
- $P(\Omega) = 1$
- Für Ereignisse $A_n, n \geq 1$, für die gilt $A_i \cap A_j = \emptyset \; \forall i \neq j$, gilt:

$$P(\bigcup_{n=1}^\infty A_n) = \sum_{n=1}^\infty P(A_n) \quad (\sigma\text{-Additivität})$$
Folgerungen aus der Definition:
- $P(\emptyset) = 0$
- $P(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n P(A_k)$ **mit** $A_k \in \mathcal{F}, \boldsymbol{A_i \cap A_j} = \emptyset \forall i \neq j$
- $A \in \mathcal{F}, P(A^c) = 1 - P(A)$
- $A \subset B$, $a, b, \in \mathcal{F}$, dann $P(A) \leq P(B)$
- stetigkeit blabla
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B \setminus A)$

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \frac{1}{n!}(n-k)(n-k-1)\ldots 1 = \frac{(n-k)!}{n!}$$
- Inklusion/Exklusion:

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

LAPLACE MODELLE UND ZÄHLEN

Definition Laplace-Modell:
Wahrscheinlichkeitsräume $(\Omega, \mathcal{F}, P)$, so dass:
- Ω endlich
- $\mathcal{F} = \mathcal{P}(\Omega)$ ="Potenzmenge von Ω"$= \{A: A \subseteq \Omega\}$
- Alle Elementarereignisse $\omega \in \Omega$ sind so, dass:

$$p(\omega) = P(\{\omega\}) = \frac{1}{|\Omega|}$$
Folgerungen:
- $P(A) = \frac{|A|}{|\Omega|}$

Injektionen oder Variationen ohne Wiederholungen:
E, F endlich und $|E| = p \leq |F| = n$ (p Objekte auf n Plätze)
Anzahl Injektionen von E nach F ist $\frac{n!}{(n-p)!}$
Anzahl Permutationen ($E = F$) ist $n!$
Kombinationen:

Problem: p identische Objekte auf $n \geq p$ Plätze verteilen
Anzahl Kombinationen ist $\binom{n}{p} = \frac{n!}{p!(n-p)!}$ mit $0 \leq p \leq n$
Teilmengen von Potenzmengen: $|\mathcal{P}(F)| = 2^{|F|}$
Binomische Formel: $(x+y)^n = \sum_{p=0}^n \binom{n}{p} x^p y^{n-p}$, $x, x \in \mathbb{R}$
Symmetrie d. Binomialkoeff.: $\binom{n}{p} = \binom{n}{n-p}$, $0 \leq p \leq n$
Formel von Pascal: $\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}$, $1 \leq p \leq n-1$
Verteilen auf Urnen, ...

ZUFALLSVARIABLEN & VERTEILUNGSFUNKT.

Definition: Zufallsvariable
Sei $(\Omega, \mathcal{F}, P)$ eine W'keitsraum. Eine Zufallsvariable ist eine Abbildung $X: \Omega \to \mathbb{R}$, so dass für $a \in \mathbb{R}$:
$$\{\omega \in \Omega, X(\omega) \leq a\} \in \mathcal{F}$$
Diese Bedingung braucht man, damit $P[\{\omega \in \Omega, X(\omega) \leq a\}]$ definiert ist.
Notation:
Man schreibt $\{X \leq a\}$ für $\{\omega \in \Omega, X(\omega) \leq a\}$. Also $P[\{X \leq a\}]$ bedeutet $P[\{\omega \in \Omega, \; X(\omega) \leq a\}]$.

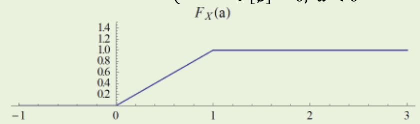
Beispiel: Sender
Sender sendet eine zufällige Nachricht von 3 Bits
Alle Folgen haben dieselbe W'keit: Laplace-Modell :
$\Omega = \{0,1\}^3 = \{\omega = (\omega_1, \omega_2, \omega_3), \omega_i \in \{0,1\}, 1 \leq i \leq 3\}$
$\omega \in \Omega \to X(\omega) = \omega_1 + \omega_2 + \omega_3 =$ „Anzahl 1 in Nachricht."
Wertebereich von $X(\omega)$ ist $\{0,1,2,3\}$

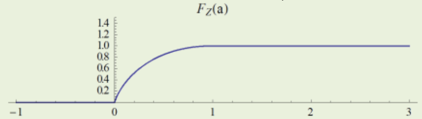
Beispiel 1: Wurf von zwei Würfeln
Wir wählen $\Omega = \{1,2,3,4,5,6\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ und $P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{36}$ für $A \subseteq \Omega$. Also ist $(\Omega, \mathcal{F}, P)$ das Laplace-Modell auf Ω.
Zufallsvariable „Augensumme der zwei Würfel":
$$\omega = (i,j) \in \Omega \xrightarrow{Z} Z(\omega) = i + j$$
Zufallsvariable „Augensummer des ersten Würfels":
$$\omega = (i,j) \in \Omega \xrightarrow{X} Z(\omega) = i$$

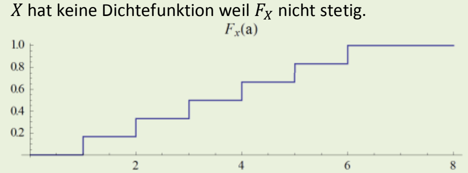
Beispiel 2: Zufälliger Punkt in $[0,1] \times [0,1]$
$\Omega = [0,1] \times [0,1]$ und $\mathcal{F}$ wie in Beispiel 1.
$P(A) =$ Fläche von $A = \int_A dx\,dy$ für $A \in \mathcal{F}$
Zufallsvariable: $X(\omega) = x$ für $\omega = (x, y)$, denn
$$\{\omega: X(\omega) \leq a\} = \begin{cases} \Omega, & a \geq 1 \\ [0, a] \times [0,1], & 0 \leq a \leq 1 \\ \emptyset, & a < 0 \end{cases}$$
und $\Omega, [0, a] \times [0,1], \emptyset$ sind Elemente von $\mathcal{F}$.
Zufallsvariable: $Y(\omega)$ analog
Zufallsvariable Fläche von $\{(0,0), (0,x), (x,y), (0,y)\}$:
$$\omega \in \Omega \xrightarrow{Z} Z(\omega) = xy, \qquad \omega = (x, y)$$
Zufallsvaribale Umfang des Rechtecks:
$$\omega \in \Omega \xrightarrow{U} U(\omega) = 2(x+y), \qquad \omega = (x, y)$$

Definition: Verteilungsfunktion
Sei X eine Zufallsvariable auf $(\Omega, \mathcal{F}, P)$. Die Verteilungsfunktion von X ist die Funktion $F: \mathbb{R} \to [0,1]$ definiert durch:
$$F(a) := P[\{X \leq a\}] = P[\{\omega \in \Omega, X(\omega) \leq a\}], \qquad a \in \mathbb{R}$$
Definition: Dichtefunktion
Eine nicht-negative $f(x)$ auf $\mathbb{R}$ heisst Dichtefunktion, wenn
$$P[\{X \leq a\}] = F(a) = \int_{-\infty}^a f(x)dx, \qquad a \in \mathbb{R}$$
Die Dichtefunktion f erfüllt die ***Normierungsbedingung***:
$$\int_{-\infty}^{\infty} f(x)dx = 1$$
Grundidentität:
$$P[\{a < X \leq b\}] = P[\{X \leq b\}] - P[\{X \leq a\}] = F(b) - F(a)$$
Somit folgt: $P[\{a < X \leq b\}] = \int_a^b f(x)dx$, $a \leq b$ in $\mathbb{R}$

Beispiel: Sender
$$F_x(0) = P(X \leq 0) = P(X = 0) = 1/8$$
$$F_x(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = 2/8, \ldots$$

Beispiel 1: Zufälliger Punkt auf $[0, 1]^2$
Zufallsvariable: $X(\omega) = x$ für $\omega = (x, y) \in [0,1]^2$, dann
$$F_X(a) = P[\{\omega: X \leq a\}] = \begin{cases} P[\Omega] = 1, \; a \geq 1 \\ P\big[[0, a] \times [0,1]\big] = a, \; 0 \leq a \leq 1 \\ P[\emptyset] = 0, \; a < 0 \end{cases}$$


X besitzt die *Dichtefunktion* $f_X(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & sonst \end{cases}$
Zufallsvariable: Fläche $Z(\omega) = xy$ für $\omega = (x, y)$
$$F_Z(a) = P(Z \leq a) - \begin{cases} 1, & a \geq 1 \\ 0, & a \leq 0 \end{cases}$$
Für $0 < a < 1$ gilt $\{Z \leq a\} = \{(x, y) \in [0,1]^2, xy \leq a\}$ und $F_Z(a) = \int_{[0,1]^2 \cap \{xy \leq a\}} dxdy = \int_0^1 \int_0^{\min\left(1, \frac{a}{x}\right)} dy\,dx = \int_0^1 \min\left(1, \frac{a}{x}\right) dx = \int_0^a dx + \int_a^1 \frac{a}{x} dx = a(1 + \log(1/a))$. Dichtefunktion f_Z:
$F_Z(a) = \int_{-\infty}^a f_Z(x)dx$ wobei $f_Z(x) = \begin{cases} \log(1/x), & 0 \leq x \leq 1 \\ 0, & sonst \end{cases}$


Beispiel 2: Wurf zweier Würfel
$X(\omega) = i$ für $\omega = (i, j) \in \{1,2,3,4,5,6\}^2 = \Omega$
$F_X(a) = P[\{X \leq a\}] = \frac{1}{36}|\{(i,j) \in \Omega, i \leq a\}| = \frac{1}{6}\big|\{i \in \{1,2,3,4,5,6, \; i \leq a\}\}\big| = \begin{cases} 0, \; a < 1 \\ k/6, \; k \leq a < k+1, \; 1 \leq k \leq 5 \\ 1, \; a > 6 \end{cases}$
X hat keine Dichtefunktion weil F_X nicht stetig.


Bedeutung der Sprunghöhe von F in k: $\frac{1}{6} = P|\{X = k\}|$

Eigenschaften der Verteilungsfunktion F:
- F wächst monoton auf $\mathbb{R}$ mit Werten in $[0,1]$
- $\lim_{a\to+\infty} F(a) = 1, \lim_{a\to-\infty} F(a) = 0$
- F ist rechtsseitig (d.h. $F(a) = \lim_{h\searrow 0} F(a+h)$)
- Für $a \in \mathbb{R}$, $P[\{X < a\}] = F(a-)$ und $P[\{X = a\}] = F(a) - F(a-)$ die Sprunghöhe von F in a".

BEDINGTE WAHRSCHEINLICHKEIT

Definition: $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, $A, B \in \mathcal{F}$ Ereignisse und $P(B) > 0$. Dann:

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

Regeln für bedingte Wahrscheinlichkeiten:
- $P(A|B) = 1 - P(A^C|B)$, weil $P(\cdot, |B)$ ein W'keitsmass
- $P(A|B) \neq 1 - P(A|B^C)$ im allgemeinen

Beispiel: Zwei Würfel

A: Augensumme $= 5$ und $\Omega = \{1,2,3,4,5,6\}^2$

$$P(A) = \frac{|\{(1,4),(2,3),(3,2),(4,1)\}|}{|\Omega|} = \frac{4}{35} = \frac{1}{9}$$

Nun wissen wir, dass erste Koord. $= 3 \to$ neuer Grundraum

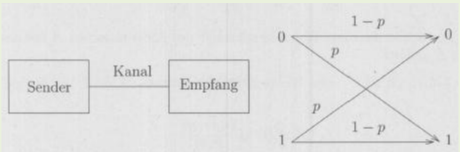
$$\Omega' = \{(3,1),(3,2),\dots,(3,6)\}$$
$$P'(A) = \frac{|\{(3,2)\}|}{|\Omega'|} = \frac{4}{36} = \frac{1}{9}$$

Formel der totalen Wahrscheinlichkeit:
$B_1, \dots, B_n \in \mathcal{F}$ Zerlegung von Ω $\left(\Omega = \bigcup_{i=1}^n B_i,\ B_i \cap B_j = \emptyset\right)$

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

Dabei sind die B_i die n möglichen Ursachen, dass A auftritt, und $P(A|B_i)$ ist die entsprechende W'keit,von A gegeben die Ursache B_i. $P(A \cap B) = P(A|B)P(B)$

Beispiel: Sender & Empfänger



Sender sendet ein Bit. Wegen Rauschen wurde das Bit mit W'keit p falsch übertragen, also mit $1-p$ richtig.
Der Sender sendet eine 0 mit W'keit π_0 und eine 1 mit W'keit $\pi_1 = 1 - \pi_0$.
X = Wert des Bits des Senders
Y = Wert des Bits am Empfang
Frage: W'keit, dass man eine 1 empfängt?

$$P(\{Y=1\}) = P(\{Y=1\}|\{X=1\})P(\{X=1\}) + P(\{Y=1\}|\{X=0\})P(\{X=0\})$$

Frage: W'keit, dass eine 1 gesendet, wenn eine 1 empfangen
$P(\{X=1\}|\{Y=1\}) = \cdots$ blabla

Formel von Bayes: $B_1, \dots B_n \in \mathcal{F}$ eine Zerlegung von Ω (d.h. $\Omega = \bigcup_{i=1}^n B_i$, $B_i \cap B_j = \emptyset\ \forall i \neq j$), mit $P(B_i) > 0$, $i = 1, \dots, n$. $A \in \mathcal{F}$ ein Ereignis mit $P(A) > 0$, dann für $i = 1, \dots, n$:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)} = \frac{P(A|B_i)P(B_i)}{P(A)}$$

$P(B_i)$: „a priori Wahrscheinlichkeit der Ursache B_i"

$P(B_i|A)$: „a posteriori W'keit der Ursache B_i (geg. A tritt auf)"

Beispiel: Sender & Empfänger

Falls man eine 1 empfängt, mit welcher W'keit wurde eine 1 gesendet (korrekte Übermittlung der Nachricht)?
Gesucht: a posteriori W'keit $P(X=1|Y=1) =$

$$\frac{P(Y=1|X=1)P(X=1)}{P(Y=1|X=0)P(X=0)+P(Y=1|X=1)P(X=1)} = \frac{(1-p)\pi_1}{p\pi_0 + (1-p)\pi_1}$$

Beispiel: Krankheitsentdeckung

Test bei Krankheit, die bei 0.1% der Population auftritt.
Bei kranken Patienten ist der Test zu 99% positiv.
Bei gesunden Patienten ist der Test zu 2% positiv.
$\{K=1\}$: Patient krank, $\{K=0\}$: Patient gesund
$\{T=1\}$: Test positiv, $\{T=0\}$: Test negativ
W'keit dass der Test mit der Ursache Patient krank:

$$P(K=1|T=1) = \frac{P(T=1|K=1)P(K=1)}{P(T=1|K=1)P(K=1)+P(T=1|K=0)P(K=0)}$$

in diesem Fall: $P(K=1|T=1) = 0.047$
Wenn Test positiv, W'keit dass krank trotzdem nur 4.7%
blabla S. 26

Beispiel: Kugeln aus Urne ziehen

Eine Urne enthält 3 schwarze und 2 weisse Kugeln.
Eine andere Urne enthält 2 schwarze und 3 weisse Kugel.
Man zieht sukzessive mit zurücklegen 2 Kugeln aus derselben, zufällig gewählten, Urne.
- W'keit, dass die zweite Kugel schwarz?
- W'keit, dass zweite Kugel schwarz, wenn erste schwarz?

S_1: Die erste gezogene Kugel ist schwarz
S_2: Die zweite gezogene Kugel ist schwarz
U_1: Die Urne mit 3 schwarzen Kugeln wurde gewählt.
U_2: Die Urne mit 2 schwarzen Kugeln wurde gewählt.
Erste Frage: $P(S_2) = ?$, zweite Frage: $P(S_2|S_1) = ?$
Es gilt: $P(S_2|U_1) = \frac{3}{5}$ und $P(S_2|U_2) = \frac{2}{5}$
Also: $P(S_2) = P(S_2|U_1)P(U_1) + P(S_2|U_2)P(U_2) = \frac{3}{5}\frac{1}{2} + \frac{2}{5}\frac{1}{2}$
Jetzt gilt: $P(S_2|S_1) = \frac{P(S_2 \cap S_1)}{P(S_1)} = 2P(S_2 \cap S_1)$
Es gilt: $P(S_2 \cap S_1|U_1) = \frac{3}{5}\frac{3}{5}$ und $P(S_2 \cap S_1|U_2) = \frac{2}{5}\frac{2}{5}$
Folglich: $P(S_2|S_1) = \left(\frac{3}{5}\frac{3}{5}\frac{1}{2} + \frac{2}{5}\frac{2}{5}\frac{1}{2}\right)/\frac{1}{2} = \frac{13}{25} > P(S_2) = \frac{1}{2}$

UNABHÄNGIGKEIT

Falls $P(A|B) = P(A)$, so hat das Auftreten von B keinen Einfluss auf das Auftreten von A. Mit der Definiton $P(A|B) = \frac{P(A \cap B)}{P(B)}$ kann man die Bedingung $P(A|B) = P(A)$ in der symmetrischen Form schreiben: $P(A \cap B) = P(A)P(B)$
Insbesondere gilt für A, B Ereignisse mit $P(A) > 0, P(B) > 0$:

$$\underbrace{P(A|B) = P(A)}_{\substack{\text{B hat keinen Einfluss}\\ \text{auf das Auftreten von A}}} \Leftrightarrow \underbrace{P(B|A) = P(B)}_{\substack{\text{A hat keinen Einfluss}\\ \text{auf das Auftreten von B}}} \Leftrightarrow P(A \cap B) = P(A)P(B)$$

Definition Unabhängigkeit: Sei $(\Omega, \mathcal{F}, P)$ ein W'keitsraum. Zwei Ereignisse $A, B \in \mathcal{F}$ (mit möglicherweise null W'keit) heissen *unabhängig*, falls $P(A \cap B) = P(A)P(B)$.
Bemerkung: Falls $0 < P(A) < 1$, so sind A und A^C *nie unabhängig*: $P(A \cap A^C) = 0 \neq P(A)P(A^C)P(A)(1 - P(A))$

Beispiel: Ziehen mit zurücklegen

Schachtel mit r schwarzen und s weissen Kugeln.
Man zieht sukzessiv n Kugeln mit Zurücklegen.
Schw. Kugeln: $1,2,\dots,r$, rote Kugeln: $r+1, r+2, \dots, r+s$
Laplace-Modell auf $\Omega = \{1,2,3,\dots,r+s\}^n$.
Für $\omega = (X_1, \dots, x_n)$ entspricht x_k der Nummer der k-ten gezogenen Kugel. Für $1 \leq k \leq n$, def. man die Zufallsvar Y_k:

$$Y_k(\omega) = \begin{cases} 1, & x_k \leq r, \qquad \omega = (x_1, \dots, x_n) \\ 0, & x_k \geq r+1 \end{cases}$$

$\{Y_k = 1\}$: das Ereignis „die k-te gezogene Kugel ist schwarz"
Zeigen, dass $\{Y_1 = 1\}$ und $\{Y_2 = 1\}$ unabhängig:
- $P[Y_1 = 1] = \frac{r(r+s)^{n-1}}{(r+s)^n} = \frac{r}{r+s}$
- $P[Y_2 = 1] = \frac{(r+s)r(r+s)^{n-2}}{(r+s)^n} = \frac{r}{r+s}$

$P[Y_1 = 1, Y_2 = 1] = \frac{r^2 r(r+s)^{n-2}}{(r+s)^n} = \frac{r^2}{(r+s)^2} = P[Y_1 = 1]P[Y_2 = 1]$
$\to \{Y_1 = 1\}$ und $\{Y_2 = 1\}$ sind unabhängig

Im Falle einer Kollektion $A_1, \dots, A_n$ von Ereignissen (d.h. $A_i \in \mathcal{F}, i = 1, .., n$) die **Kollektion unabhängig**, falls $\forall 1 \leq i_1 < i_2 < \cdots < i_k \leq n,\ 2 \leq k \leq n$ gilt:

$$P[A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}] = P[A_{i_1}] \dots P[A_{i_k}]$$

Beispiel

$\Omega = \{1,2,3,4\}, \quad A = \{1,2\}, \quad B = \{2,3\}, \quad C = \{1,3\}$
$P[A \cap B] = \frac{1}{4} = P[A]P[B], P[A \cap C] = \frac{1}{4} = P[A]P[C],$
$P[B \cap C] = \frac{1}{4} = P[B]P[C]$
Aber: $P[A \cap B \cap C] = 0 \neq P[A]P[B]P[C]$
Also hat Eintreten von $B \cap C$ Einfluss auf A!

Man definiert $Y_1, Y_2, Y_3, Y_4: \Omega \to \{0,1\}$

$Y_i = \mathbb{1}_{S_i} = \begin{cases} 1, & falls\ \omega \in S_i \\ 0, & sonst \end{cases}$, so dass $\{Y_i = 1\} = S_i\ \forall i$

heissen **Bernoulli Variablen** mit Erfolgsparameter $P[Y_i = 1]$.

Beispiel

Unabhängige $0-1$-experimente mit Erfolgsparameter p, $0 \leq p \leq 1$.

Unabhängige 0-1-Experimente mit Erfolgsparam. $\boldsymbol{p} \in [\mathbf{0}, \mathbf{1}]$:
Jetzt $\Omega_{(n)} = \{0,1\}^n$ „0-1-Folgen der Länge n"
Wir setzen $Q_p(\{\omega\}) = p^{\#1\ in\ \omega}(1-p)^{\#0\ in\ \omega}$.

Kein Laplace-Modell: $Q(\{(1,1,1,1)\}) = \left(\frac{2}{5}\right)^4 \neq Q(\{(1,0,1,1)\})$

blabla s.30 im skript

DISKRETE VERTEILUNGEN

DEFINITION

Eine **Zufallsvariable** $X: \Omega \to \mathbb{R}$ auf $(\Omega, \mathcal{F}, P)$ heisst **diskret**, falls die Menge $E = \{x \in \mathbb{R}, \{X = x\} \neq \emptyset\}$ (d.h. Wertebereich von X) höchstens abzählbar ist.
Verteilung der diskreten Zufallsvariablen X:

$$p(x) = P[\{X = x\}], \qquad x \in E$$

Die Verteilung $p(x)$ erfüllt die ***Normierungsbedingung***:

$$\sum_{x \in E} p(x) = 1, \qquad p(x) \geq 0, \qquad x \in E$$

Bernoulliverteilung: $X: \Omega \to \{0,1\}$

$$p = P[\{X = 1\}], \qquad 1 - p = P[\{X = 0\}]$$

mit Erfolgsparameter $p \in [0,1]$
Binomialverteilung:
Wir betrachten eine unabh. Folge von 0-1-Experimenten mit Erfolgsparameter p der Länge n. (wie letztes Beispiel in Kap. 1)

$$S_n(\omega) := X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega), \qquad \omega \in \{0,1\}^n$$

die „totale Anzahl der Erfolge in der Folge ω"
Also $0 \leq S_n(\omega) \leq n$ und für $0 \leq k \leq n$:

$P_p[\{S_n = k\}] = P_p[\{\omega = (\epsilon_1, \dots, \epsilon_n), \#\{i, \epsilon_i = 1\} = k\}] = \sum_{\substack{\omega = (\epsilon_1, \dots, \epsilon_n) \\ mit\ \#\{i, \epsilon_i = 1\} = k}} P_p[\{\omega\}] = p^k(1-p)^{n-k} \cdot \#\{\omega \in \{0,1\}^n,$
$\omega\ mit\ k\ "1"\ und\ (n-k)\ "0"\} = \binom{n}{k} p^k (1-p)^{n-k}$

Die **Verteilung der Zufallsvariablen** S_n („total Anzahl der Erfolge") ist die **Binomialverteilung** der Länge n mit Erfolgsparameter p und:

$$P[S_n = k] := b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad 0 \leq k \leq n$$

Beispiel

Eine Schachtel enthält 1 schwarze und 2 weisse Kugeln. Man zieht 12 Kugeln mit zurücklegen. Was ist die W'keit, dass man genou 4 schwarze Kugeln zieht?

$$P[genau\ 4\ schwarze\ Kugeln] = b_{12,1/3}(4) = \binom{12}{4}$$

blabla

balbla

GEOMETRISCHE VERTEILUNG

Stelle ersten Erfolgs in einer unendl. Folge von 0-1-Experimenten mit Erfolgsfaktor p.
„kein Gedächtnis" für die Verteilung.

Beispiel Münzwurf

Man wirft wiederholt ein nicht faire Münze. Auftreten von Kopf ist p. Die n_0 ersten Auftreten zeigen Zahl. Was ist W'keit, dass das erste Auftreten von Kopf k Würfe später kommt? Zufallsvariable T:
T: geom. zufallsvert. (p): Stelle von ersten Auftreten von Kopf: wir suchen $P[T = k + n_0|T > n_0]$

$$P[T = k] = p(1-p)^{k-1}, \qquad k \geq 1$$

und: $P[T = k + n_0|T > n_0] = P[T = k] \to$ kein Gedächtnis

POISSON-VERTEILUNG

Approximation der Binomialverteilung $b_{n,p}(k)$ für $n \to \infty$ und $n\,p \approx \lambda, \lambda > 0$ eine feste Zahl. Es handelt sich um die „Verteilung der totalen Anzahl Erfolge bei n unabh. Versuchen, die alle eine kleine Erfolgschance $\left(\approx \frac{\lambda}{n}\right)$ haben".

$$p_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k \geq 0, \qquad \lambda > 0$$

und heisst **Poissonverteilung mit Parameter** λ.
Diese Approximation ist sehr nützlich in Modellen, in denen viele unabhängige Ursachen mit kleiner W'keit auftreten, z.B. Anzahl Druckfehler auf einer Seite eines Buches, die totale Anz. Telefonanrufe zu einer gew. Zeit in einer Telefonzentrale

ERWARTUNGSWERT FÜR DISKRETE VARIABLEN

Definition Erwartungswert einer diskreten Variablen:

$$E[X] = \sum_{x \in E_X} x\, p_X(x), \qquad falls \sum_{x \in E_X} |x| p_X(x) < \infty$$

mit $p_X(x) = P[X = x]$ die Verteilung von X.
Eigenschaften des Erwartungswertes:
Sei X eine diskr. Zufallsvariable mit Verteilung $p_X(x)$, $g: \mathbb{R} \to \mathbb{R}$ eine Funktion. Wir können die Zufallsvariable $Y = g(X) (= g \circ X)$ betrachten. Um $E[Y]$ berechnen zu können,

müssten wir definitionsgemäss die Verteilung von Y berechnen. Es gilt aber:

$$E[Y] = \sum_{x\in E_X} g(x)p_X(x), \text{ falls } \sum_{x\in E_X}|g(x)|p_X(x) < \infty$$

und die Bedingung $\sum_{x\in E_X}|g(x)|p_X(x) < \infty$ ist äquivalent zu $\sum_{y\in E_Y}|y|p_Y(y) < \infty$, denn: $p_Y(y) = \sum_{x:g(x)=y} p_X(x)$, und folglich:

$$\sum_y |y|p_Y(y) = \sum_y |y| \sum_{x:g(x)=y} p_X(x)$$
$$= \sum_y \sum_{x:g(x)=y}|g(x)|p_X(x) = \sum_x |g(x)|p_X(x)$$

Daraus folgt $\sum_{y\in E_Y}|y|p_Y(y) < \infty \Leftrightarrow \sum_{x\in E_X}|g(x)|p_X(x) < \infty$.

blablabla

Beispiel

$T\sim\text{geom}(p)$ mit $p > 0$, dann: $g: t \mapsto t^2$

$E[T^2] = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = \sum_{k=1}^{\infty} \underbrace{k(k-1)}_{(=0\, für\, k=1)} p(1-p)^{k-1} + \underbrace{\sum_{k=1}^{\infty} kp(1-p)^{k-1}}_{=E[T]=\frac{1}{n}} = p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} + \frac{1}{p} = p(1-p)\left(\sum_{k=0}^{\infty} z^k\right)_{z=1-p}^{n} + \frac{1}{p} = p(1-p)\left[\frac{1}{1-z}\right]_{z=1-p}^{n} + \frac{1}{p} = p(1-p)\frac{2}{(1-(1-p))^3} + \frac{1}{p} = \frac{2-p}{p^2}$

Linearität:

X, Y Zufallsvar. mit $\sum_{x\in E_X}|x|p_X(x) < \infty, \sum_{y\in E_Y}|y|p_Y(y) < \infty$

$$E[\lambda X + \mu Y] = \lambda E[X] + \mu E[Y], \qquad \lambda, \mu \in \mathbb{R}$$

Dazu führen wir die **gemeinesame Verteilung von X und Y** ein:

$$p_{X,Y}(x,y) = P[X = x, Y = y], \qquad x \in E_X, \qquad y \in E_Y$$

Die gemeinsame Verteilung erfüllt:

$$p_X(x) = \sum_{y\in E_Y} p_{X,Y}(x,y), \qquad x \in E_X$$

dasselbe für $p_Y(y)$ analog.

Beispiel

Eine Urne, 2 weisse, 3 schwarze, 1 blaue Kugel
2 Kugeln werden gezogen mit zurücklegen.
X = Anzahl gez. weisse Kugeln, Y = Anz. gez. blaue Kugeln

Y \ X	0	1	2
0	1/4	1/3	1/9
1	1/6	1/9	0
2	1/36	0	0

Zeilensummen: $p_Y(y)$, Spaltensummen: $p_X(x)$

Unabhängigkeit: Falls $p_{X,Y}(x,y) = p_X(x)p_Y(y)\forall x \in E_X, y \in E_Y$
Dann gilt: $E[XY] = E[X]E[Y]$

Varianz $\sigma_X^2 = \text{var}(X)$ der diskreten Zufallsvariablen X:

$$\sigma_X^2 = E[(X-m)^2] = E[X^2] - E[X]^2, \qquad m = E[X]$$
$$\sigma_X^2 = \sum_{x\in E_X}(x-m)^2 p_X(x)$$

Varianz ist Kennzahl für Grösse d. Fluktuationen von X um m

Standardabweichung: $\sigma_X = \sqrt{\sigma_X^2}$

Eigenschaften der Varianz:
- $\sigma_{ax+b}^2 = a^2\sigma_X^2$
- $X_1, \ldots$ paarw. unabh. und $S = X_1 + \cdots \Rightarrow \sigma_S^2 = \sigma_{X_1}^2 + \cdots + \sigma_{X_n}^2$

Beispiel: Varianz der geom. Verteilung

$$T\sim\text{geom}(p), \sigma_T^2 = E[T^2] - E[T]^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$

Im Allg. $\text{var}(X_1 + X_2) = \text{var}(X_1) + 2\cos(X_1, X_2) + \text{var}(X_2)$
mit $\cos(X_1, X_2) = E[(X_1 - m_1)(X_2 - m_2)]$

Covarianz: blablabla

S.53 - 56

Anmerkung Normalvertielung.

$$X_1, X_2 \sim N(0,1), \qquad Y = \alpha X_1 + \beta X_2$$
$$\rightarrow Y \sim N(0, \sigma^2 = \alpha^2 + \beta^2)$$

Jatzt falls X_1, X_2 unabh. $N(m_1, \sigma_1)$ bzw $N(m_2, \sigma_2)$ verteilt, was ist die Verteilung von $Y = \alpha X_1 + \beta X_2$?

$X_1 = \sigma_1 \underbrace{\left(\frac{X_1 - m_1}{\sigma_1}\right)}_{\tilde{X}_1} + m_1$, X_2 analog.

Nun sind $\tilde{X}_1, \tilde{X}_2 \sim N(0,1)$-verteilt

$$m = E[Y] = \alpha E[X_1] + \beta E[X_2] = \alpha m_1 + \beta m_2$$
$$\sigma^2 = var[Y] = var[\alpha X_1 + \beta X_2] = \cdots = \alpha^2\sigma_1^2 + \beta^2\sigma_2^2$$

balbla

ÜBERSICHT VERTEILUNGEN

BERNOULLI-VERTEILUNG

W'keits-Funktion: $P[X = k] = \begin{cases} p, & k = 1 \\ 1-p, & k = 0 \end{cases}$

Verteilungsfunkt.: $P[X \le k] = \begin{cases} 0, & k < 0 \\ 1-p, & 0 \le k < 1 \\ 1, & k \ge 1 \end{cases}$

bzw. $P[X < k] = \begin{cases} 0, & k \le 0 \\ 1-p, & 0 < k \le 1 \\ 1, & k > 1 \end{cases}$

Erwartungswert: $E(X) = p$
Varianz: $\text{Var}(X) = p(1-p)$

BINOMIAL-VERTEILUNG

Verteilungsfunkt: $F_X(x) = \sum_{k=0}^{\lfloor x\rfloor}\binom{n}{k}p^k(1-p)^{n-k}$

Erwartungswert: $E(X) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = np$

Varianz:
$\text{Var}(X) = \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} - n^2p^2 = np(1-p) = npq$

Zusammenhang Normaleverteilung:
Binomialverteilung wird mit $n \to \infty$ zur Normalverteilung.

NEGATIVE BINOMIALVERTEILUNG

$$P[X = n] = \binom{n-1}{r-1}p^r(1-p)^{n-r}$$

Erwartungswert: $E(X) = r/p$
Varianz: $\text{Var}(X) = r(1-p)/p^2$
Spezialfall: geom. Verteilung für $r = 1$

POISSON-VERTEILUNG

Verteilungsfunkt.: $F_\lambda(n) = \sum_{k=0}^{n} P_\lambda(k) = e^{-\lambda}\sum_{k=0}^{n}\frac{\lambda^k}{k!}$

Erwartungswert: $E(X) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda$

Varianz: $\text{Var}(X) = \sum_{k=0}^{\infty}(k-\lambda)^2\frac{\lambda^k}{k!}e^{-\lambda} = \lambda$

ergänzen mit Zusammenfassung

GAMMAVERTEILUNG

Dichtefunktion: $f(x) = \begin{cases} \frac{b^p}{\Gamma(p)}x^{p-1}e^{-bx}, & x \ge 0 \\ 0, & x < 0 \end{cases}$

Verteilungsfunkt.: $F(x) = \begin{cases} P(p, bx), & x \ge 0 \\ 0, & x < 0 \end{cases}$

Erwartungswert: $E(X) = p/b$
Varianz: $\text{Var}(X) = p/b^2$

NORMALVERTEILUNG (GAUSS-VERTEILUNG)

Verwendung: Messfehler, Brown'sche Bewegung
Interpretation: Zentrierungsparam: $m = \mu$, Breite: σ
Spezialfall: Standardnormalerverteilung: $\sigma = 1$

Dichtefunktion: $f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$

Verteilungsfunkt.: $F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(\frac{-1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right)dt$

Erwartungswert: $E(X) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}x\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)dx = \mu$

Varianz:

$$\text{Var}(X) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\mu)^2\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)dx = \sigma^2$$

Standardabw.: $\sigma = \sqrt{\text{Var}(X)}$

INVERSE GAUSS-VERTEILUNG

Dichtefunktion: $f(x) = \begin{cases} \sqrt{\frac{\lambda}{2\pi x^3}}\exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right), & x > 0 \\ 0, & x \le 0 \end{cases}$

Erwartungswert: $E(X) = \mu$
Varianz: $\text{Var}(X) = \mu^3/\lambda$
Standardabw.: $\sigma = \sqrt{\text{Var}(X)} = \sqrt{\mu^3/\lambda}$

EXPONENTIALVERTEILUNG

Verwendung: radioaktiver Zerfall, Lebensdauer von Bauteilen
Interpretation: λ: Ausfallrate, $\frac{1}{\lambda}$: Lebensdauer

Dichtefunktion: $f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$

Verteilungsfunkt.: $F(x) = \int_{-\infty}^{x} f_\lambda(t)dt = \begin{cases} 1 - e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$

Median: $\tilde{x} = \ln(2)/\lambda$
Erwartungswert: $E(X) = \int_0^{\infty}\lambda x e^{-\lambda x}dx = 1/\lambda$

Varianz: $\text{Var}(X) = \int_0^{\infty}\left(x - \frac{1}{\lambda}\right)^2\lambda e^{-\lambda x}dx = 1/\lambda^2$

Standardabw.: $\sigma = \sqrt{\text{Var}(X)} = \sqrt{1/\lambda^2} = 1/\lambda$

INTEGRALE

$\int_0^{\infty} t^n e^{-t}dt = n!$

Erwartungswert kontinuierlicher Variablen:
falls nur Dichte von X bekannt:

$$E[Y] = \int_{-\infty}^{\infty} g(x)f_X(x)dx, \qquad Y = g(X)$$

balbla

Unabhängigkeit ist nichts automatisches:

Bsp

W'keit Panne 1 Generator an einem Tag ist 10^{-3}.
4 Gen. Ausfall am einem Tag: $(10^{-3})^4$ unabhängig!
Solche Situation erscheint nach N Jahren wobei $N\frac{365}{10^{12}} \approx 1$
Missbrauch von Unabhängigkeit: Kausalität wurde nicht beachtet! Blitz → alle 4 Generatoren defekt

Allgemeine Tipps zur Vorlesung:
- versuchen via Komplement zu rechnen
- A_i unabhängig $\Rightarrow P(A_1 \cap A_2) = P(A_1)P(A_2)$
- De-Morgan: $A \cap B = (A^C \cup B^C)^C$, $A \cup B = (A^C \cap B^C)^C$

Definition: $A\Delta B := (A \cup B) \cap (A \cap B)^C$

ÜBUNGSSTUNDEN

SERIE 2

W'keitsraum: $(\Omega, \mathcal{F}, P)$
- Ω: Grundraum
- $\mathcal{F}$: σ-Algebra, wenn Ω endl., meistens $\mathcal{F} = \mathcal{P}(\Omega)$
- P: W'keitsmass, $P: \mathcal{F} \to [0,1] = P(\Omega)$
  - $P(\Omega) = 1$
  - $A_n \in \mathcal{F},\ A_n \cap A_m = \emptyset\ \forall n \ne m$
    $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

blablabla

ADMINISTRATIVES

Prüfung:
90 min., 5 Blätter doppelseitig handbeschrieben. Kein Rechner.

TODO:
- Prinzip Inklusion/Exklusion
- verallgem. W'keit beim Roboter
- tipps & aufgabenst. aus Übungen allgemein

STATISTIK

- Testen von statistischen Hypothesen
- Schätzen von Parametern