① Warscheinligkeitsraum

→ Warscheinligkeitsraum ist ein Trippel $(\Omega, \mathcal{F}, \mathcal{P})$

- Ω = Grundraum
- $\mathcal{F}$ = Ereignisse → Menge aller Teilmengen von Ω
- $\mathcal{P}$ = Funktion $A \in \mathcal{F} \to P(A)$

→ Eigenschaften:

i. $P(\Omega) = 1$
ii. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
iii. $\forall A_n,\ n \geq 1$, disjunkt gilt $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$

② Verteilungsfunktionen

→ Zufallsvariable X: Abbildung $X: \Omega \to \mathbb{R}$ so, dass für $a \in \mathbb{R}$

$\{\omega \in \Omega,\ X(\omega) \leq a\} \in \mathcal{F} \longrightarrow P[\{\omega \in \Omega,\ X(\omega) \leq a\}]$

→ Verteilungsfunktion: Falls es eine Funktion $f(x)$ existiert mit $P[\{X \leq a\}] = F(a) = \int_{-\infty}^{a} f(x)\,dx$, so heisst $f(x)$ die Dichtefunktion von X.

→ Eigenschaften:

i. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

ii. F wächst monoton auf $\mathbb{R}$ mit Werten in $[0,1]$

iii. $\lim_{a\to\infty} F(a) = 1,\quad \lim_{a\to-\infty} F(a) = 0$

iv. $P[X < a] = F(a-)$

v. $P[X = a] = F(a) - F(a-)$

vi. $P[a < X \leq b] = F(b) - F(a) = \int_a^b f(x)\,dx$

Falls es überhaupt eine Dichtefunktion existiert

③ Bedingte Warscheinligkeit

→ Bedingte Warscheinligkeit von A gegeben $B \Rightarrow P(A|B)$

→ Totale Warscheinligkeit $\quad P(A) = \sum_{i=0} P(A|B_i)\,P(B_i)$

→ Bayes:

$$\boxed{P(B|A) = \frac{P(B\cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A|B)P(B)}{\sum_{i=0} P(A|B_i)P(B_i)}}$$

Totale Warscheinligkeit

→ Unabhängigkeit:

i. $P(A|B) = P(A)$ → auftreten von B kein Einfluss auf A

ii. $P(A\cap B) = P(A)\,P(B)$

iii. A und A^c sind nie unabhängig
Auftreten von A hat Einfluss auf A^c

④ Diskrete Verteilungen

1. Bernoulli Verteilung

$X: \Omega \to \{0,1\}$, p = Erfolgsparameter $(0 \leq p \leq 1)$

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

↳ man schaft es oder man schaft es nicht

2. Binomialverteilung

→ Anzahl Erfolge in n Versuche → Warscheinligkeit, dass man in k von n Versuche erfolgreich ist.

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

3. Geometrische Verteilung

→ Erster Erfolg in einer Folge von n unabhängige Experimente

$$P[T^n = k] = g_p(k) = p(1-p)^{k-1}$$

4. Poisson Verteilung

→ Approximation von Binomialverteilung $b_{n,p}(k)$ für $n \to \infty$ $(np \sim \lambda)$

Verteilung der totalen Anzahl der Erfolge bei zahlreichen (=n) unabhängigen Versuchen, die alle eine kleine Erfolgschance ($\sim \lambda/n$) haben.

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

5 Erwartungswert → Diskrete Verteilungen

→ Definition: Eine Art von Mittelwert

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

→ Eigenschaften

i. X und Y unabhängig $\Rightarrow E[XY] = E[X]E[Y]$

ii. $X, Y \rightarrow E[\lambda X + \mu Y] = \lambda E[X] + \mu E[Y]$

iii. Für X, $p_X(x)$ und $g: \mathbb{R} \to \mathbb{R}$ mit $Y = g(X)$

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

Beispiel: $Y = X^2$ und X geometrisch verteilt mit p.

$$\Rightarrow E[Y] = E[X^2] = \sum_{x \in E_X} g(x) p (1-p)^{x-1} = \sum_{k=1}^{\infty} k^2 p (1-p)^{k-1} = \frac{2-p}{p^2}$$

$g(x) = x^2 \Rightarrow g(x) = x^2$

iv. Für grosses n gilt $E[X] \approx \sum_i^n \frac{x_i}{n}$ (empirische Mittel)

Varianz

→ $\sigma_X^2 = Var[X] := E[(X-m)^2]$ mit $m = E[X]$

$$\Rightarrow \sigma_X^2 = E[X^2] - E[X]^2$$

→ Standartabweichung: $\sigma_X = \sqrt{Var[X]}$

→ Eigenschaften

i. Falls $X_1, X_2, \dots, X_n$ paarweise unabhängige Zufallsvariablen sind und $S = X_1 + X_2 + \dots + X_n$

$$\Rightarrow \sigma_S^2 = \sum_{i=1}^n E[(X_i - m_i)^2] = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2 = \sum_{i=1}^n \sigma_i^2$$

ii. $X, Y \rightarrow Var[\lambda X + \mu Y] = \lambda^2 Var[X] + \mu^2 E[Y]$

6 Stetige Verteilungen

1. Gleichverteilung

$$f_{a,b} = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{sonst} \end{cases} \longrightarrow F_{a,b}(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x \geq b \end{cases}$$

2. Exponentialverteilung

→ Analog zur geometrischen Verteilung

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{sonst} \end{cases} \longrightarrow F_\lambda(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}$$

- $E[X] = 1/\lambda$
- $Var[X] = 1/\lambda^2$

3. Normalverteilung

→ Approximation der Binomialverteilung

- m = Zentrierungsparameter ($m \in \mathbb{R}$)
- σ = Breiteparameter ($\sigma > 0$)

$$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

→ Satz von de Moivre-Laplace

$$P[a \leq \tilde{S}_n \leq b] \xrightarrow{n\to\infty} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a)$$

wobei $\Phi(t) := \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ für $t \in \mathbb{R}$

→ Standartnormal: $m = 0$, $\sigma = 1$

- $E[X] = 0$, $Var[X] = 1$

→ Nicht standartnormal: für $m \neq 0$, $\sigma \neq 1$

Angenommen Y ist standartnormal und X normalverteilt mit Parameter m und σ

$$\Rightarrow X = \sigma Y + m \rightarrow P[X \leq a] = P[\sigma Y + m \leq a] = P[Y \leq \tfrac{a-m}{\sigma}]$$

- $E[X] = m$, $Var[X] = \sigma^2$

7 Erwartungswert → Stetige Verteilungen

→ Definition: Eine Art von Mittelwert

$$\boxed{E[X] = \int_{-\infty}^{\infty} x f(x)\, dx, \quad \text{falls} \int_{-\infty}^{\infty} |x| f(x)\, dx < \infty}$$

↳ $f(x)$ = Dichtefunktion von X

→ Eigenschaften:

i. X und Y unabhängig $\Rightarrow E[XY] = E[X]E[Y]$

ii. $X, Y \rightarrow E[\lambda X + \mu Y] = \lambda E[X] + \mu E[Y]$

iii. Für X, $p_X(x)$ und $g: \mathbb{R} \to \mathbb{R}$ mit $Y = g(X)$

$$\Rightarrow E[Y] = \int_{-\infty}^{\infty} g(x) p_X(x)\, dx, \quad \text{falls} \int_{-\infty}^{\infty} |g(x)| p_X(x)\, dx < \infty$$

Varianz

$$\rightarrow \sigma_X^2 = Var[X] := E[(X-m)^2] = \int_{-\infty}^{\infty} (x-m)^2 f(x)\, dx \quad \text{mit } m = E[X]$$

$$\Rightarrow \sigma_X^2 = E[X^2] - E[X]^2$$

→ Standartabweichung: $\sigma_X = \sqrt{Var[X]}$

→ Eigenschaften

i. Für $Y = aX + b$, $a, b \in \mathbb{R} \Rightarrow Var[Y] = a^2 Var[X]$

8 Gemeinsame Dichte

→ X, Y Zufallsvariablen: $\Omega \to \mathbb{R}$ auf $(\Omega, \mathcal{F}, P)$ besitzen eine gemeinsame stetige Verteilung, falls $f(x,y) \geq 0$ auf $\mathbb{R}^2$ existiert, so dass

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

„gemeinsame Dichte von (X,Y)" ↵

→ Falls X und Y eine gemeinsame Dichte besitzen

$$F_X(a) = P[X \leq a] = P[X \leq a, -\infty \leq Y \leq \infty] = \int_{-\infty}^{a}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy\, dx$$

$$\Rightarrow F_X(a) = \int_{-\infty}^{a} \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \right) dx$$

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

→ Eigenschaften

i. X und Y unabhängig $\Rightarrow f_{X,Y}(x,y) = f_X(x) f_Y(y)$

ii. X und Y unabhängig $\Rightarrow E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

iii. X und Y unabhängig $\Rightarrow Var[X+Y] = Var[X] + Var[Y]$

↳ im allgemein ist $Var[X+Y] = Var[X] + Var[Y] + 2cov(X,Y)$

→ Kovarianz

$$cov(X,Y) := E[(X-m_x)(Y-m_y)] \quad m_x = E[X], \; m_y = E[Y]$$

$$cov(X,Y) = E[XY] - E[X]E[Y]$$

- X, Y independent $\Rightarrow cov(X,Y) = 0$

→ Korrelationskoeffizient

$$\rho(X,Y) := \frac{cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \in [-1, 1]$$

→ **Beste Lineare Prognose**: Man will X möglichst gut durch Zufallsvariable $\hat{X}$ der Form $\hat{X} = \alpha Y + \beta$ mit $\alpha, \beta \in \mathbb{R}$ vorhersagen. Prognose von X durch Y ist gegeben durch

$$\hat{X} = \frac{cov(X,Y)}{Var[Y]}(Y - E[Y]) + E[X]$$

9 Statistik

1. Maximum-Likelihood Schätzer

→ Poisson Verteilung: $p_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \geq 0 \quad (\lambda > 0)$

- Wir betrachten n unabhängie Poisson Verteilungen und definieren

$$f_n(\lambda, x_1, x_2, \dots) = \lambda^n e^{-\lambda(x_1 + x_2 + \dots)}$$

$\Rightarrow$ Wähle λ so, dass f_n maximiert wird

$$\frac{\partial \log(f_n)}{\partial \lambda} = 0 \rightarrow \log(f_n) = n\log(\lambda) - \lambda \sum_i^n x_i \Rightarrow \boxed{\tilde{\lambda} = \frac{n}{\sum_i^n x_i}}$$

→ Normalverteilung: $f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$

$$f_n(m, \sigma, x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n}\frac{1}{\sigma^n}\exp\left(-\sum_1^n \frac{(x_i-m)^2}{2\sigma^2}\right), \quad m \in \mathbb{R}, \sigma > 0$$

- $\frac{\partial}{\partial m}\log(f_n) = 0 = -\frac{1}{\sigma^2}\sum_1^n (x_i - m) \Rightarrow \boxed{\tilde{m} = \frac{1}{n}\sum_i^n x_i}$
- $\frac{\partial}{\partial \sigma}\log(f_n) = 0 = \frac{n}{\sigma} - \frac{1}{\sigma^3}\sum_i^n (x_i - m)^2 \Rightarrow \boxed{\tilde{\sigma}^2 = \frac{1}{n}\sum_i^n (x_i - \tilde{m})^2}$

→ Laplace Verteilung: für $x_1, x_2, \dots, x_n = 0$ oder 1 (success / fail)

- um p zu schätzen $\Rightarrow \boxed{\tilde{p} = \frac{1}{n}\sum_i^n x_i}$

→ Gleichverteilung: Gleichverteilung auf $[a,b]$

- Da nur Werte zwischen a und b erlaubt sind

$$\Rightarrow \boxed{\tilde{a} = \min_{1 \leq i \leq n} x_i, \; \tilde{b} = \max_{1 \leq i \leq n} x_i}$$

Stetige Verteilung $\Rightarrow f_n(x_1, x_2, \dots, x_n) = \prod_i^n f_{X_i}(X_i)$

→ Geometrische Verteilung: $P[T=k] = p(1-p)^k$

$$f_n(p, x_1, x_2, \dots, x_n) = \prod_i^n p(1-p)^{T_i - 1} = p^n (1-p)^{\left(\sum_i^n T_i - 1\right)}$$

$$\log(f_n) = n\log(p) + \left[\sum_i^n T_i - 1\right]\log(1-p)$$

$$\log(f_n) = n\log(p) + \left[\sum_i^n T_i - 1\right] \log(1-p)$$

maximum at $\frac{d}{dp}\log(f_n) = 0$

$$\Rightarrow \frac{n}{p} = \frac{1}{1-p}\sum_i^n T_i - 1 \Rightarrow \frac{n}{p} = \left(\sum_i T_i\right) \cancel{-n} \cancel{+n} \Rightarrow \boxed{p = \frac{n}{\sum_i^n T_i}}$$

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

$$L(p; x_1, x_2, \dots) = \prod_i^m \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}$$

m = # repetions / data
n = # of possibilities for the distribution

$$\log(L(p; x_1, x_2, \dots)) = \sum_i^m \binom{n}{x_i} + x_i \log(p) + (n-x_i)\log(1-p)$$

$$\frac{d}{dp} L(p; x_1, x_2, \dots) = \sum_i^m \frac{x_i}{p} + \frac{x_i - n}{1-p} \stackrel{!}{=} 0 \Rightarrow \sum_i^m x_i =: S_m$$

$$\frac{1}{p} S_m + \frac{1}{1-p} S_m - \frac{mn}{1-p} \stackrel{!}{=} 0 \rightarrow \frac{1-p}{p} S_m = mn - S_m$$

$$\Rightarrow \boxed{\tilde{p} = \frac{S_m}{m \cdot n} = \frac{\sum_i^m x_i}{m \cdot n}}$$

→ General Maximum-Likelihood: The probability to observe the sequence $x_1, x_2, \dots, x_n$

$$\boxed{L(x_1, x_2, \dots) = \prod_i^n p(x_i)}$$

→ Natural estimator:

$$\boxed{\mathbb{E}[X] \approx \sum_i^m \frac{x_i}{m} \quad \text{for large } n}$$

Example:

i. Binomial distribution: $\mathbb{E}[X] = n \cdot p \approx \sum_i^m \frac{x_i}{m} \Rightarrow \tilde{p} \approx \sum_i^n \frac{x_i}{m \cdot n}$

ii. Poisson: $\mathbb{E}[X] = \lambda \approx \sum_i^m \frac{x_i}{m} \Rightarrow \tilde{\lambda} \approx \sum_i^m \frac{x_i}{m}$

	$E[X]$	$Var[X]$	Dichtefunktion	Verteilungsfunktion
Bernoulli Verteilung	p	$p(1-p)$		
Binomialverteilung	np	$np(1-p)$		
Geometrische Verteilung	$\frac{1}{p}$	$\frac{1}{p^2}-\frac{1}{p}$		
Poisson Verteilung	λ	λ		
Gleichverteilung	$\frac{a+b}{2}$	$\frac{1}{12}(b-a)^2$	$f_{a,b} = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{sonst} \end{cases}$	$F_{a,b}(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x \geq b \end{cases}$
Exponentialverteilung	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{sonst} \end{cases}$	$F_\lambda(x) = \begin{cases} 0, & x \leq 0 \\ 1-e^{-\lambda x}, & x > 0 \end{cases}$
Normalverteilung	m	σ^2	$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$	$F_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2} dt$

$$\rightarrow \sum_{k=1}^{\infty} k x^{k-1} = \frac{d}{dx}\left(\sum_{k=1}^{\infty} x^k\right) = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}$$