



# 64-544

## Grundlagen der Signalverarbeitung und Robotik

[http://tams.informatik.uni-hamburg.de/  
lectures/2014ss/vorlesung/GdSR](http://tams.informatik.uni-hamburg.de/lectures/2014ss/vorlesung/GdSR)

Jianwei Zhang, Bernd Schütz



Universität Hamburg  
Fakultät für Mathematik, Informatik und Naturwissenschaften  
Fachbereich Informatik  
**Technische Aspekte Multimodaler Systeme**

Sommersemester 2014

# Gliederung

1. Einführung
2. Grundlagen der Robotik
3. Grundlagen der Sensorik
4. Verarbeitung von Scandaten
5. Rekursive Zustandsschätzung
6. Fuzzy-Logik
7. Steuerungsarchitekturen





# Agenda

## 5. Rekursive Zustandsschätzung

Grundlagen der Wahrscheinlichkeitsrechnung

Stochastische Fortpflanzungsgesetze und Belief

Bayes-Filter

Selbst-Lokalisierung mobiler Roboter

Literatur





# Rekursive Zustandsschätzung

- ▶ **Idee:** Schätzen eines Systemzustands, der nicht direkt gemessen werden kann, aber aus den Messungen ableitbar ist
  - ▶ Beobachtungen bestimmter Aspekte des Systems
  - ▶ Steueraktionen, die das System beeinflussen
  - ▶ Unsicherheit in Beobachtungen und Aktionen
- ▶ probabilistische Algorithmen zur Zustandsschätzungen berechnen eine **belief distribution** über die möglichen Zustände
- ▶ der **belief** beschreibt das Wissen (Schätzung) eines Systems über seinen Zustand
- ▶ Sensorwerte, Stellgrößen, der Zustand eines Systems und von dessen Umgebung können als **Zufallsvariable** modelliert werden



# Grundlagen der Wahrscheinlichkeitsrechnung

- ▶ ein **Zufallsexperiment** ist ein grundsätzlich beliebig oft wiederholbarer Vorgang mit den Eigenschaften:
  - ▶ reproduzierbarer Aufbau und Durchführung
  - ▶ alle möglichen **Ergebnisse** ( $\omega_i$ ) sind vorab bekannt (endliche Anzahl)
  - ▶ Nichtvorhersagbarkeit, Zufälligkeit des Ergebnisses
- ▶ Menge  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  heißt **Ergebnisraum** eines Zufallsexperiments
  - ▶ jedem Experiment ist höchstens ein Element  $\omega_i \in \Omega$  zugeordnet
- ▶ jede Teilmenge  $A$  eines Ergebnisraumes  $\Omega$  heißt **Ereignis**
  - ▶ ein Ereignis  $A$  tritt ein, wenn ein Ergebnis  $\omega_i$  vorliegt, mit  $\omega_i \in A$
  - ▶ einelementige Teilmengen (also die Ereignisse, die genau ein Ergebnis  $\omega$  enthalten) werden als Elementarereignisse bezeichnet
  - ▶  $\mathcal{P}(\Omega)$  ist die Menge aller Teilmengen des Ergebnisraumes  $\Omega$  (bei Abzählbarkeit von  $\Omega$  die Potenzmenge) und heißt **Ereignisraum**



# Grundlagen der Wahrscheinlichkeitsrechnung

## Beispiel: Würfel

Zufallsexperiment: Augenzahl beim Wurf eines Würfels

- ▶ Ergebnisraum:  $\Omega = \{1, 2, 3, 4, 5, 6\}$
- ▶ Ereignis: Teilmenge  $A \subseteq \Omega$ 
  - ▶ alle geraden Augenzahlen:  $A = \{2, 4, 6\}$
  - ▶ Augenzahlen  $< 4$ :  $B = \{1, 2, 3\}$
  - ▶ Augenzahlen  $> 6$ :  $C = \{\}$
  - ▶ Elementarereignisse:  $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$
- ▶ Ereignisraum:  $\mathcal{P}(\Omega) =$   
 $\{\{\}, \{1\}, \{1, 2\}, \{1, 2, 3\}, \dots, \{1, 2, 3, 4, 5, 6\}, \{2\}, \{2, 3\}, \dots, \{5, 6\}, \{6\}\}$



## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

- ▶ Eine Funktion  $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ , die jedem Ereignis  $A \in \mathcal{P}(\Omega)$  eine reelle Zahl  $P(A)$  zuordnet, heißt **Wahrscheinlichkeitsmaß**.

Diese Funktion muss folgende Eigenschaften erfüllen:

1.  $P(A) \geq 0$  (Axiom I) (12)

2.  $P(\Omega) = 1$  (Axiom II) (13)

3.  $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$  (Axiom III) (14)

(Axiome nach Kolmogorow)

- ▶ ein mögliches Wahrscheinlichkeitsmaß  $P$ :  
wird jedem Ergebnis aus  $\Omega$  (Elementarereignis) die gleiche Wahrscheinlichkeit zugeordnet, so gilt:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\# \text{ der Ergebnisse } \in A}{\# \text{ aller möglichen Ergebnisse}}$$

(Laplace'scher Wahrscheinlichkeitsbegriff)



# Grundlagen der Wahrscheinlichkeitsrechnung

## Beispiel: Würfel

Zufallsexperiment: Augenzahl beim Wurf eines Würfels

- ▶ Ereignisraum:  $\mathcal{P}(\Omega) =$   
 $\{\{\}, \{1\}, \{1, 2\}, \{1, 2, 3\}, \dots, \{1, 2, 3, 4, 5, 6\}, \{2\}, \{2, 3\} \dots \{5, 6\}, \{6\}\}$
- ▶ Wahrscheinlichkeitsmaß  $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$   
 $P(\{\}) = 0, P(\{1\}) = \frac{1}{6}, P(\{1, 2\}) = \frac{2}{6}, P(\{1, 2, 3\}) = \frac{3}{6}, \dots$   
 $\dots, P(\{1, 2, 3, 4, 5, 6\}) = 1, P(\{2\}) = \frac{1}{6}, P(\{2, 3\}) = \frac{2}{6}, \dots$   
 $\dots, P(\{5, 6\}) = \frac{2}{6}, P(\{6\}) = \frac{1}{6}$
- ▶ Wahrscheinlichkeitsmaß nach Laplace, z.B.:  
 $A = \{\omega \mid \omega_{\text{geradeZahl}}\} = \{2, 4, 6\}$   
 $P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$



## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

► Folgerungen aus den Axiomen:

- $P(\emptyset) = 0$
- $P(A) = 1 - P(\bar{A})$ , mit  $\bar{A}$  Komplementärereignis zu  $A$
- $B \subseteq A \Rightarrow P(B) \leq P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A_1 \cup A_2 \cup \dots \cup A_m) = P(A_1) + P(A_2) + \dots + P(A_m)$   
wenn die Paare  $A_i, A_j$  mit  $i \neq j$  disjunkt sind
- $P(A \setminus B) = P(A) - P(A \cap B)$

Beispiel:

$$A = \{\omega \mid \omega \text{ geradeZahl}\} = \{2, 4, 6\} \text{ und } B = \{\omega \mid \omega > 3\} = \{4, 5, 6\}$$

$$P(A \setminus B) = P(\{2, 4, 6\} \setminus \{4, 5, 6\}) = P(\{2\}) = \frac{1}{6}$$

$$P(A \setminus B) = P(A) - P(A \cap B) = P(\{2, 4, 6\}) - P(\{2, 4, 6\} \cap \{4, 5, 6\}) = P(\{2, 4, 6\}) - P(\{4, 6\}) = \frac{3}{6} - \frac{2}{6} = \frac{1}{6}$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Zufallsvariable

- ▶ sei  $\Omega$  ein diskreter Ergebnisraum, dann bezeichnet eine Abbildung

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

eine reelle **Zufallsvariable**  $X$

- ▶ wegen  $\Omega$  diskret ist der Wertebereich  $W = X(\Omega)$  von  $X$  ebenfalls diskret mit  $W = \{x_1, x_2, \dots, x_m\}$
- ▶ die Wahrscheinlichkeiten der Elementarereignisse in  $\Omega$  ( $P(\omega)$ ) induzieren Wahrscheinlichkeiten auf die Ereignisse  $x_1, x_2, \dots$  in  $W$ :

$$P_X(\{x_i\}) = P(\{X = x_i\}) = P(\{\omega \in \Omega | X(\omega) = x_i\})$$

Wahrscheinlichkeit, dass  $X$  den Wert  $x_i$  hat

- ▶ verkürzter Schreibweise:

$$p(x_i) = P_X(x_i) = P(X = x_i)$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Wahrscheinlichkeitsverteilung und Wahrscheinlichkeitsfunktion

- ▶ Wahrscheinlichkeit, dass Zufallsvariable  $X$  den Wert  $x_i$  hat:

$$p(x_i) = P_X(x_i) = P(X = x_i)$$

- ▶  $P_X$  heißt **Wahrscheinlichkeitsverteilung** von  $X$  und wird hier durch die Wahrscheinlichkeitsfunktion  $p$  dargestellt
- ▶ die **Wahrscheinlichkeitsfunktion**  $p: \mathbb{R} \rightarrow [0, 1]$  einer diskreten Zufallsvariablen  $X$  ist für  $x \in \mathbb{R}$  definiert durch:

$$p(x) = \begin{cases} p_i = P(X = x_i), & x = x_i \in \{x_1, x_2, \dots, x_n\} \\ 0, & \text{sonst} \end{cases} \quad (15)$$

und es gilt  $\sum_i p_i = 1$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Zufallsvariable, Beispiel

Zufallsexperiment: Augenzahl beim Wurf zweier Würfel

▶ Ergebnisraum:

$$\Omega = \{(1|1), (1|2), \dots, (1|6), (2|1), (2|2), \dots, (6|6)\}$$

$$|\Omega| = 36$$

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

- ▶  $X$  hier Abbildung der Ereignisse auf Augenzahl
- ▶ Wertebereich  $W = \{2, 3, 4, 5, 6, \dots, 11, 12\}$
- ▶  $X((1|1)) = 2, X((1|2)) = 3, \dots, X((1|6)) = 7, \dots$   
 $\dots, X((2|1)) = 3, X((2|2)) = 4, \dots, X((6|6)) = 12$
- ▶  $\{X = 5\} = \{\omega \mid X(\omega) = 5\} = \{(1|4), (2|3), (3|2), (4|1)\}$
- ▶  $p(5) = P_x(5) = P(X = 5) = \frac{4}{36}$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Zufallsvariable, Beispiel (cont.)

- ▶ Wahrscheinlichkeitsverteilung repräsentiert durch Wahrscheinlichkeitsfunktion  $p : \mathbb{R} \rightarrow [0, 1]$

$$p(x) = \begin{cases} \frac{1}{36} & x = 2 \\ \frac{2}{36} & x = 3 \\ \frac{3}{36} & x = 4 \\ \vdots & \vdots \\ \frac{6}{36} & x = 7 \\ \vdots & \vdots \\ \frac{2}{36} & x = 11 \\ \frac{1}{36} & x = 12 \\ 0 & \text{sonst} \end{cases}$$



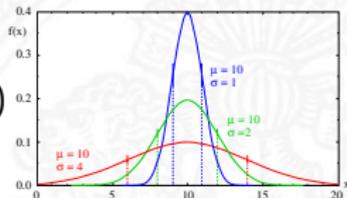


# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Dichtefunktion

- ▶ nimmt die Zufallsvariable kontinuierliche Werte an, wird sie durch eine **Dichtefunktion**  $f : \mathbb{R} \rightarrow [0, \infty[$ , oder auch Wahrscheinlichkeitsdichte, beschrieben (engl. *probability density function*, *PDF*)
- ▶ eine typische Dichtefunktion ist die **Normalverteilung** mit Mittelwert  $\mu$  und Varianz  $\sigma^2$ :

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right\} \quad (16)$$



- ▶ Funktionswert einer PDF ist nicht wie bei diskreten Wahrscheinlichkeiten nach oben durch 1 begrenzt
- ▶ für eine PDF gilt:

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



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Verteilungsfunktion

- ▶ die **Verteilungsfunktion**  $F: \mathbb{R} \rightarrow [0; 1]$  einer Zufallsvariablen  $X$  ist für  $x \in \mathbb{R}$  definiert durch:

$$F(x) = P(X \leq x) \quad (17)$$

- ▶  $F(x)$  berechnet die Wahrscheinlichkeit, dass der Wert der Zufallsvariablen  $X \leq$  vorgegebenem Wert  $x$  ist.
- ▶ es gelten die Eigenschaften:
  - ▶  $F$  ist monoton steigend
  - ▶  $F$  ist rechtsseitig stetig
  - ▶  $0 \leq F(x) \leq 1$ ;  $\left( \lim_{x \rightarrow -\infty} F(x) = 0; \lim_{x \rightarrow \infty} F(x) = 1 \right)$
- ▶ unter Anwendung der Axiome nach Kolmororow (siehe: 12, 13, 14) gilt:

$$P(a < X \leq b) = F(b) - F(a)$$

# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Verteilungsfunktion

- ▶ nach Kolmogorow (siehe: 12, 13, 14) gilt:

$$P(a < X \leq b) = F(b) - F(a)$$

die Mengen (Ereignisse)  $\{X \leq a\} = \{\omega | X(\omega) \leq a\}$ ,  
 $\{a < X \leq b\} = \{\omega | a < X(\omega) \leq b\}$  und  
 $\{X > b\} = \{\omega | X(\omega) > b\}$

sind paarweise disjunkt:

$$1 \stackrel{(13)}{=} P(\Omega) \stackrel{(14)}{=} P(X \leq a) + P(a < X \leq b) + P(X > b)$$

$$\begin{aligned} P(a < X \leq b) &= 1 - P(X \leq a) - P(X > b) & | & P(X \leq b) = 1 - P(X > b) \\ &= P(X \leq b) - P(X \leq a) & | & F(x) = P(X \leq x) \\ &= F(b) - F(a) \end{aligned}$$

# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

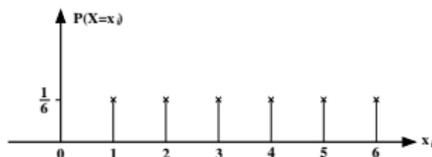
## diskrete Verteilungsfunktion

- ▶ Eine Funktion  $F$ , die durch die diskrete Zufallsvariable  $X$  bestimmt ist, heißt diskrete **Verteilungsfunktion** von  $X$ .

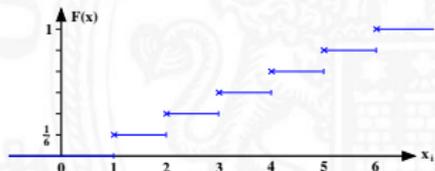
$$F(x) = P(X \leq x), \quad x \in \mathbb{R} \quad (\text{siehe (17)})$$

$$F(x) = P(X \leq x) = \sum_{x_j \leq x} P(X = x_j) = \sum_{x_j \leq x} p(x_j)$$

- ▶  $F(x)$  ist im diskreten Fall eine stückweise konstante Treppenfunktion, mit  $P(X = x)$  als Sprunghöhe im Punkt  $x$
- ▶ Beispiel Würfel:



Wahrscheinlichkeitsfkt



Verteilungsfunktion



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

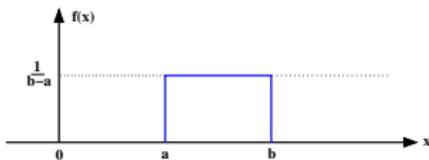
## kontinuierliche Verteilungsfunktion

- ▶ Eine Funktion  $F$ , die durch die kontinuierliche Zufallsvariable  $X$  bestimmt ist, heißt kontinuierliche **Verteilungsfunktion** von  $X$ .

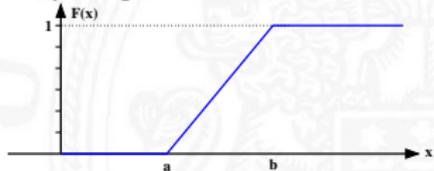
$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du \quad (18)$$

mit  $f : \mathbb{R} \rightarrow [0, \infty[$  als Dichtefunktion von  $X$

- ▶ Beispiel Gleichverteilung im Intervall  $(a, b]$ :



Dichtefunktion



Verteilungsfunktion

# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## kontinuierliche Verteilungsfunktion & Dichtefunktion

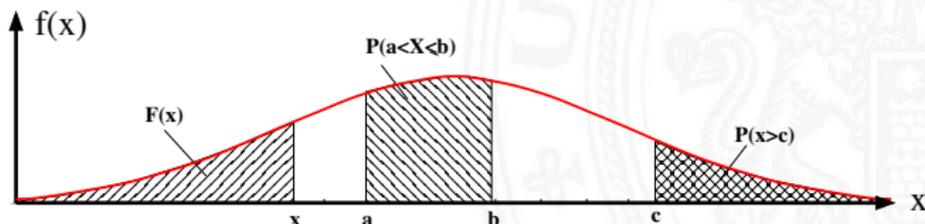
- ▶ Berechnung der stetigen Verteilung durch ihre Dichte  $f(x)$ :

$$F(u) = \int_{-\infty}^u f(x) dx \quad (18)$$

Wert der stetigen Zufallsvariablen im Intervall  $(a, b]$ :

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x) dx$$

- ▶ Beispiel: Dichte einer stetigen Zufallsvariablen  $X$





# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Erwartungswert

- ▶ Der **Erwartungswert** einer Zufallsvariablen ist gegeben als:

$$E[X] = \mu = \sum_x x p(x) \quad (\text{diskret})$$

$$E[X] = \mu = \int x f(x) dx \quad (\text{kontinuierlich})$$

- ▶ Der Erwartungswert ist eine lineare Funktion einer Zufallsvariablen, d.h. es gilt:

$$E[aX + b] = aE[X] + b$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Varianz

- ▶ Die **Varianz**  $\sigma^2$  von  $X$  berechnet sich wie folgt:

$$\sigma^2 = E[(X - \mu)^2] = E[X^2] - \mu^2$$

- ▶ Die Varianz misst das erwartete Quadrat der Abweichung vom Erwartungswert
- ▶ Für die Standardabweichung oder Streuung  $\sigma$  von  $X$  gilt:

$$\sigma = \sqrt{\sigma^2} = \sqrt{E[(X - \mu)^2]}$$



# Grundlagen der Wahrscheinlichkeitsrechnung

## Erwartungstert u. Varianz am Beispiel

Zufallsexperiment: Augenzahl beim Wurf eines Würfels

| 7-Würfel  
(0..6)

- ▶ Erwartungswert  $E[X] = \mu = \sum_x x p(x)$

$$\mu = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = \frac{1}{6} * 21$$

$$= 3,5$$

$$= 3,0$$

- ▶ Varianz  $\sigma^2 = E[(X - \mu)^2] = E[X^2] - \mu^2$ :

$$\sigma^2 = (1 - 3,5)^2 * \frac{1}{6} + (2 - 3,5)^2 * \frac{1}{6} + (3 - 3,5)^2 * \frac{1}{6}$$

$$+ (4 - 3,5)^2 * \frac{1}{6} + (5 - 3,5)^2 * \frac{1}{6} + (6 - 3,5)^2 * \frac{1}{6}$$

$$\approx 2,92$$

$$= 4$$

- ▶ Standardabweichung  $\sigma = \sqrt{\sigma^2}$ :

$$\sigma \approx 1,71$$

$$= 2$$

# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Erwartungstert u. Varianz am Beispiel

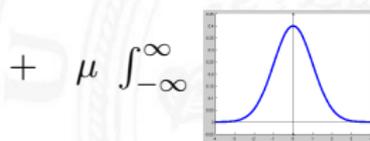
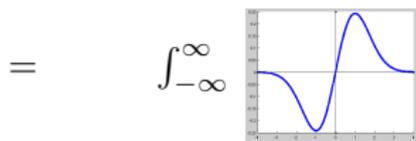
Erwartungswert  $E[X] = \mu = \int x f(x) dx$  der Normalverteilung

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

Substitution:  $t = x - \mu$ ;  $x = t + \mu$ ;  $\frac{dt}{dx} = 1$ ;  $dx = dt$ ;

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (t + \mu) \exp\left(-\frac{1}{2}\frac{t^2}{\sigma^2}\right) dt$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} t \exp\left(-\frac{1}{2}\frac{t^2}{\sigma^2}\right) dt + \mu \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{t^2}{\sigma^2}\right) dt$$



$$= \mu$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## mehrere Zufallsvariablen

- ▶ Für die Wahrscheinlichkeit, dass Zufallsvariable  $X$  den Wert  $x$  und Zufallsvariable  $Y$  den Wert  $y$  hat, mit  $p$  als Wahrscheinlichkeitsverteilung, schreibt man:

$$p(x, y) = P(X = x \text{ und } Y = y)$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Kovarianz

- ▶ Für die **Kovarianz** (Cov) zweier reeller Zufallsvariablen  $X, Y$  gilt:

$$\text{Cov}(X, Y) = E((X - E(X)) \cdot (Y - E(Y)))$$

unter der Voraussetzung, dass  $E(X)$ ,  $E(Y)$  und  $E(XY)$  existiert.

- ▶  $\text{Cov}(X, Y) > 0$ , monoton wachsender linearer Zusammenhang
- ▶  $\text{Cov}(X, Y) = 0$ , falls kein linearer Zusammenhang zwischen den Variablen; möglicherweise aber ein nichtlinearer Zusammenhang (Variablen sind nicht notwendigerweise stochastisch unabhängig)
- ▶  $\text{Cov}(X, Y) < 0$ , monoton fallender linearer Zusammenhang
- ▶ keine Aussage über die Stärke des Zusammenhangs
- ▶  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- ▶  $\text{Cov}(X, X) = \text{Var}(X) = \sigma^2$
- ▶ falls  $X, Y$  stochastisch unabhängig, folgt  $\text{Cov}(X, Y) = 0$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Dichtefunktionen, mehrdimensional

- ▶ im Falle mehrerer reeller Zufallsvariablen werden die paarweisen Kovarianzen als Matrix notiert
- ▶ die Kovarianzmatrix  $\Sigma$  ist die Matrix der paarweisen Kovarianzen eines Zufallsvektors  $X = (X_1, \dots, X_n)$

$$\Sigma = \begin{bmatrix} \text{Cov}(X_1, X_1) & \cdots & \text{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \cdots & \text{Cov}(X_n, X_n) \end{bmatrix}$$

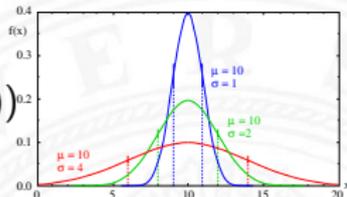
- ▶ Die **Kovarianzmatrix**  $\Sigma$  ist *positiv semidefinit* und *symmetrisch* ( $xAx^T \geq 0$ , diagonalisierbar, Eigenwerte  $\geq 0$ )

# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Kovarianzmatrix und mehrdimensionale Dichtefunktion

- ▶ eine Normalverteilung mit Mittelwert  $\mu = 10$  und Varianz  $\sigma^2 = 1, 2, 4$ :

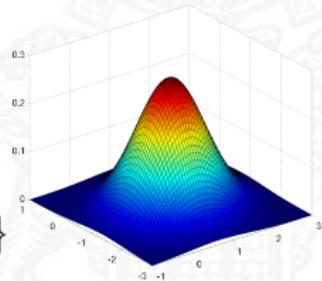
$$p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \quad (\text{siehe: (16)})$$



- ▶ im mehrdimensionalen Fall mit  $x = [a \ b]$  und Mittelwertsvektor  $\mu = [1 \ -1]$  und

$$\text{Kovarianzmatrix } \Sigma = \begin{bmatrix} 0.6 & 0.1 \\ 0.1 & 0.6 \end{bmatrix}:$$

$$p(x) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x - \mu)\Sigma^{-1}(x - \mu)^T\right\}$$





# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Stochastische Unabhängigkeit

- ▶ Für die Wahrscheinlichkeit, dass  $X$  den Wert  $x$  und  $Y$  den Wert  $y$  hat, mit  $p$  als Wahrscheinlichkeitsverteilung, schreibt man

$$p(x, y) = P(X = x \text{ und } Y = y)$$

- ▶ Zwei Ereignisse  $x$  und  $y$  heißen **stochastisch unabhängig** (unter  $p$ ), wenn gilt:

$$p(x, y) = p(x) \cdot p(y) \quad (19)$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## bedingte Wahrscheinlichkeiten

- ▶ Wahrscheinlichkeit für  $P(X = x)$  unter der Voraussetzung  $P(Y = y)$  (**bedingte Wahrscheinlichkeit**):

$$p(x|y) = P(X = x|Y = y)$$

- ▶ wenn  $p(y) > 0$ , so gilt: 
$$p(x|y) = \frac{p(x, y)}{p(y)} \quad (20)$$

- ▶ sind  $X$  und  $Y$  von einander unabhängig, gilt (siehe Gl. 19):

$$p(x|y) = \frac{p(x)p(y)}{p(y)} = p(x)$$

→ wenn  $X$  und  $Y$  unabhängig von einander sind, sagt  $Y$  nichts über  $X$  aus



# Grundlagen der Wahrscheinlichkeitsrechnung

## stochastische Unabhängigkeit am Beispiel des Würfels

▶ Zufallsexperiment 1:

beim aufeinanderfolgenden Wurf zweier Würfel fällt zuerst Ereignis  $A$  (Augenzahl 1 oder 4) und dann Ereignis  $B$  (ebenfalls Augenzahl 1 oder 4):

$$P(A, B) = P(A) \cdot P(B) = \frac{2}{6} \cdot \frac{2}{6} = \frac{1}{9}$$

▶ Zufallsexperiment 2:

beim aufeinanderfolgenden Wurf zweier Würfel treten in beliebiger Reihenfolge eine 1 und eine 4 auf

→ Ereignis  $A = \{1, 4\}$ ; Ereignis  $B = \{1, 4\} \setminus \{\omega_1\}$

$$P(A, B) = P(A) \cdot P(B|A) = \frac{2}{6} \cdot \frac{1}{6} = \frac{1}{18}$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## bedingte Unabhängigkeit und bedingte Unabhängigkeit

Zwei Zufallsvariablen  $X$  und  $Y$  heißen

- ▶ **unabhängig**, falls gilt:

$$p(x, y) = p(x) \cdot p(y) \quad (\text{siehe Gl. 19})$$

- ▶ **bedingt unabhängig**, falls gilt:

$$p(x, y | z) = p(x|z) \cdot p(y|z) \quad (21)$$

mit  $Z$  als weitere Zufallsvariable

- ▶  $Y$  trägt keine Information über  $X$ , wenn  $Z$  bekannt ist



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## totale Wahrscheinlichkeit

- Gesetz der totalen Wahrscheinlichkeit:

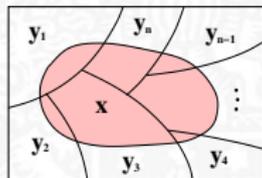
$$p(x) = \sum_y p(x|y)p(y) \quad (\text{diskret})$$

$$p(x) = \int p(x|y)p(y)dy \quad (\text{kontinuierlich})$$

Seien  $y_1, y_2, \dots, y_n$  wechselweise disjunkte Ereignisse, die den Ereignisraum  $S$  bilden, also  $S = (y_1 \cup y_2 \cup y_3 \cup \dots \cup y_n)$ ; ferner liege das Ereignis  $x$  ebenfalls in  $S$

$$\begin{aligned} x &= S \cap x = (y_1 \cup y_2 \cup y_3 \cup \dots \cup y_n) \cap x \\ &= (y_1 \cap x) \cup (y_2 \cap x) \cup \dots \cup (y_n \cap x) \quad \text{mit (14)} \end{aligned}$$

$$\begin{aligned} p(x) &= p(y_1, x) + p(y_2, x) + \dots + p(y_n, x) \\ &= p(x|y_1)p(y_1) + p(x|y_2)p(y_2) \dots p(x|y_n)p(y_n) \\ &= \sum_{i=1}^n p(x|y_i)p(y_i) \end{aligned}$$





## Exkurs: Wahrscheinlichkeitsbegriff

- ▶ objektivistische Wahrscheinlichkeit  
Wahrscheinlichkeiten basieren direkt auf beobachtbaren Naturvorgängen und lassen sich numerisch beziffern
  - ▶ Frequentismus  
Wahrscheinlichkeit wird als Grenzwert der relativen Häufigkeit eines Ereignisses interpretiert.
- ▶ subjektivistische Wahrscheinlichkeit  
Wahrscheinlichkeit als Maß für die Sicherheit der persönlichen Einschätzung
  - ▶ Thomas Bayes (1702 – 1761)  
Wahrscheinlichkeit: “degree of belief”;  
A-priori-Wahrscheinlichkeit: Vorwissen (persönliches) und Annahmen;  
A-posteriori-Wahrscheinlichkeit ermittelt sich aus a-priori-Wahrscheinlichkeit plus zusätzlicher Beobachtung eines Ereignisses



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Satz von Bayes

- ▶ **Satz von Bayes:** (mit  $p(y) > 0$ )

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\sum_{x'} p(y|x')p(x')} \quad (\text{diskret})$$

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y|x')p(x')dx} \quad (\text{kontinuierlich})$$

Satz von Bayes beschreibt Umkehrung von Schlussfolgerungen

- ▶ die Berechnung von  $p(\text{Ereignis}|\text{Ursache})$  ist oft einfach
- ▶ aber gesucht ist  $p(\text{Ursache}|\text{Ereignis})$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Satz von Bayes: Beispiel

In einer Zahnradfabrik werden Zahnräder identischer Abmessungen auf zwei verschiedenen Maschinen  $m_1$  und  $m_2$  gefräst. Die Maschine  $m_1$  stellt 40% der Zahnräder her, die Maschine  $m_2$  60%. Dabei sind 4% der Zahnräder der Maschine  $m_1$  fehlerhaft und 8% der Maschine  $m_2$ .

Aus der Gesamtproduktion der Zahnräder wird eines zufällig entnommen.

Dieses ist fehlerhaft. Wie hoch ist die Wahrscheinlichkeit, dass es von Maschine  $m_1$  gefräst wurde?

gesucht:  $p(m_1|f)$

$$p(m_1) = 0.4; \quad p(m_2) = 0.6; \quad p(f|m_1) = 0.04; \quad p(f|m_2) = 0.08$$

$$\begin{aligned} p(m_1|f) &= \frac{p(f|m_1)p(m_1)}{p(f)} = \frac{p(f|m_1)p(m_1)}{\sum_{i=1}^2 p(f|m_i)p(m_i)} \\ &= \frac{0.04 \cdot 0.4}{(0.04 \cdot 0.4) + (0.08 \cdot 0.6)} = \frac{0.016}{0.064} = \frac{1}{4} \end{aligned}$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Satz von Bayes

Beweisidee:

$$\begin{aligned}
 p(x, y) &= p(y, x) && \text{(mit Gl. (20): } p(x, y) = p(x|y)p(y)\text{)} \\
 p(x|y)p(y) &= p(y|x)p(x) \\
 p(x|y) &= \frac{p(y|x)p(x)}{p(y)}
 \end{aligned}$$

Einsetzen der totalen Wahrscheinlichkeit im Nenner:

$$p(x_k|y) = \frac{p(y|x_k)p(x_k)}{\sum_{i=1}^n p(y|x_i)p(x_i)}$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Satz von Bayes

- ▶ Der Satz von Bayes ist vielfältig anwendbar.
- ▶ Sei  $x$  die Größe, die wir aus  $y$  ableiten wollen, dann bezeichnet man die Grundwahrscheinlichkeit  $p(x)$  als **A-Priori-Wahrscheinlichkeit** und  $y$  als **Daten** (z.B. Sensormessungen).
- ▶ Die Verteilung der  $p(x)$  beschreibt das Wissen über  $X$ , bevor die Messung  $y$  berücksichtigt wird.
- ▶ Die Verteilung der  $p(x|y)$  wird als **A-Posteriori-Wahrscheinlichkeit** von  $X$  bezeichnet.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Satz von Bayes

- ▶ Der Satz von Bayes erlaubt die bedingte Wahrscheinlichkeit  $p(x|y)$  zu berechnen, wenn die Wahrscheinlichkeit für das Ergebnis einer Messung  $y$  bei einem Zustand  $x$  ( $p(y|x)$ ) und die Grundwahrscheinlichkeiten  $p(x)$  und  $p(y)$  bekannt sind.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$



# Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

## Satz von Bayes

- ▶ Im Satz von Bayes hängt die Grundwahrscheinlichkeit  $p(y)$  nicht vom betrachteten  $x$  ab.
- ▶ Daher ist der Faktor  $p(y)^{-1}$  für jeden Wert  $x$  in  $p(x|y)$  gleich.
- ▶ Dieser Faktor wird meist als Normalisierungsfaktor im Satz von Bayes notiert:

$$p(x|y) = \eta p(y|x)p(x)$$

- ▶ Der Faktor  $\eta$  wird derart gewählt, dass die Normierung der Verteilung auf 1 erhalten bleibt.
- ▶  $\eta$  bezeichnet im Folgenden immer solche Normalisierungsfaktoren, deren eigentliche Werte sich aber unterscheiden können.



## Grundlagen der Wahrscheinlichkeitsrechnung

- ▶ Alle bisherigen Regeln können durch (eine) weitere Zufallsvariable(n), z. B.  $Z$ , bedingt sein.
- ▶ dies bedeutet für den Satz von Bayes mit  $Z = z$ :

$$\begin{aligned}
 p(x, y, z) &= p(y, x, z) && (p(x, y, z) = p(x|y, z)p(y, z)) \\
 p(x|y, z)p(y, z) &= p(y|x, z)p(x, z)
 \end{aligned}$$

$$p(x|y, z) = \frac{p(y|x, z)p(x|z)}{p(y|z)} \quad (22)$$

solange  $p(y|z) > 0$



## Grundlagen der Wahrscheinlichkeitsrechnung (cont.)

- falls  $p(x | y, z)$  und  $Y$  trägt keine Information über  $X$ , falls  $Z$  bekannt ist (bed. Unabhängigkeit von  $X$  und  $Y$ ), lässt sich der Ausdruck  $p(x | y, z)$  vereinfachen?

$$p(x, y | z) = p(x|z)p(y|z) \quad \text{bed. Unabhängigkeit nach (21)}$$

$$p(x|z) = \frac{p(x, y | z)}{p(y|z)}$$

$$p(x, y | z) = \frac{p(x, y, z)}{p(z)} \quad \text{bed. W-keit nach (20)}$$

$$= \frac{p(x | y, z) p(y, z)}{p(z)}$$

$$= \frac{p(x | y, z) p(y|z) p(z)}{p(z)}$$

$$p(x | y, z) = \frac{p(x, y, z)}{p(y, z)}$$

$$p(y|z) = \frac{p(y, z)}{p(z)}$$

$$p(x|z) = \frac{p(x | y, z) p(y|z)}{p(y|z)}$$

$$p(x | y, z) = p(x|z) \quad (23)$$



# Grundlagen der Wahrscheinlichkeitsrechnung

## bedingte Unabhängigkeit

- ▶  $p(x, y | z) = p(x|z)p(y|z)$  besagt, dass  $y$  keine Information über  $x$  trägt, wenn  $z$  bekannt ist

- ▶ Formel beschreibt eine **bedingte Unabhängigkeit** und ist äquivalent zu

$$p(x|z) = p(x | y, z) \quad \text{vgl. (23)}$$

$$p(y|z) = p(y | x, z)$$

- ▶ impliziert **nicht**, dass  $X$  unabhängig von  $Y$  ist:

$$p(x, y | z) = p(x|z)p(y|z) \not\Rightarrow p(x, y) = p(x)p(y)$$

die Umkehrung gilt generell auch nicht:

$$p(x, y) = p(x)p(y) \not\Rightarrow p(x, y | z) = p(x|z)p(y|z)$$



# Stochastische Fortpflanzungsgesetze

- ▶ Der Zustand  $x_t$  eines Systems lässt sich durch eine Wahrscheinlichkeitsverteilung

$$p(x_t \mid x_{0:t-1}, z_{1:t-1}, u_{1:t})$$

beschreiben, die abhängig ist von

- ▶ den bisherigen Zuständen  $x_{0:t-1}$
- ▶ allen bisherigen Messungen  $z_{1:t-1}$  und
- ▶ allen bisherigen Stellgrößen (Steuerkommandos)  $u_{1:t}$



## Stochastische Fortpflanzungsgesetze (cont.)

- ▶ ist der Zustand  $x_t$  **komplett**, fasst die Stellgröße  $u_t$  und der Zustand  $x_{t-1}$  alle bisherigen Ereignisse zusammen:

$$p(x_t \mid x_{0:t-1}, z_{1:t-1}, u_{1:t}) = p(x_t \mid x_{t-1}, u_t) \quad (24)$$

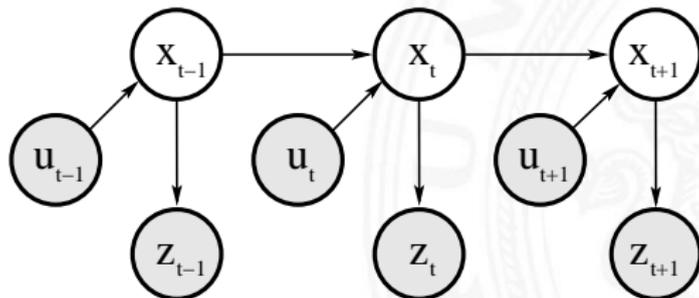
- ▶ für die Messungen gilt dann entsprechend:

$$p(z_t \mid x_{0:t}, z_{1:t-1}, u_{1:t}) = p(z_t \mid x_t) \quad (25)$$

- ▶ der Zustand  $x_t$  ist ausreichend, um die Messung  $z_t$  vorherzusagen
- ▶ die bedingte Wahrscheinlichkeit  $p(x_t \mid x_{t-1}, u_t)$  wird als **Zustandsübergangswahrscheinlichkeit** bezeichnet
  - ▶ Änderung des Zustandes abhängig von den Stellgröße
- ▶ die Wahrscheinlichkeit  $p(z_t \mid x_t)$  wird **Wahrscheinlichkeit der Messung** genannt

## Stochastische Fortpflanzungsgesetze (cont.)

- ▶ Zustandsübergangswahrscheinlichkeit und Wahrscheinlichkeit der Messung beschreiben zusammen ein **dynamisches stochastisches System**
- ▶ eine solche Systembeschreibung ist auch bekannt als **hidden Markov model (HMM)** oder **dynamic Bayes network (DBN)**





## Zustandsschätzung: Belief

- ▶ Als **belief** bezeichnet man das Wissen (Schätzung) eines Systems über seinen Zustand.
- ▶ Der *wahre Zustand* eines Systems ist nicht dasselbe!
- ▶ Der *belief* ist die A-Posteriori-Wahrscheinlichkeit der Zustandsvariablen anhand der bisherigen Daten:

$$\text{bel}(x_t) = p(x_t \mid z_{1:t}, u_{1:t}) \quad (26)$$

- ▶ diese Definition berechnet den *belief* (als Wahrscheinlichkeit) **nach** der aktuellen Messung  $z_t$ .



## Zustandsschätzung: Belief (cont.)

- ▶ Der *belief vor* der Messung wird als

$$\overline{bel}(x_t) = p(x_t \mid z_{1:t-1}, u_{1:t}) \quad (27)$$

definiert und meist als **Vorhersage** (engl. *prediction*) bezeichnet.

- ▶ der Schritt zur Berechnung von  $bel(x_t)$  aus der Vorhersage  $\overline{bel}(x_t)$  wird als **Korrektur** (engl. *correction* oder *measurement update*) bezeichnet



# Bayes-Filter

- ▶ Ein allgemeiner Algorithmus, um *beliefs* zu berechnen, ist der **Bayes-Filter-Algorithmus**.
- ▶ der Algorithmus ist rekursiv und berechnet  $bel(x_t)$  zum Zeitpunkt  $t$  aus folgenden Größen
  - ▶  $bel(x_{t-1})$  zum Zeitpunkt  $t - 1$  (alter belief)
  - ▶ der Stellgröße  $u_t$  (aktuelle Stellgröße)
  - ▶ der Messung  $z_t$  (aktuelle Messung)



# Bayes-Filter-Algorithmus

Bayes-Filter ( $bel(x_{t-1}), u_t, z_t$ ):

1. **for all**  $x_t$  **do**
2.      $\overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1}$
3.      $bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t)$
4. **endfor**
5. **return**  $bel(x_t)$



## Bayes-Filter-Algorithmus (cont.)

- ▶ in Zeile 2 wird die Stellgröße  $u_t$  verarbeitet:
  - ▶  $\overline{bel}(x_t)$  ist das Integral des Produktes zweier Wahrscheinlichkeitsverteilungen:
    - ▶ die Wahrscheinlichkeit für Zustand  $x_{t-1}$  und
    - ▶ die Wahrscheinlichkeit, dass mit  $u_t$  in den Zustand  $x_t$  gewechselt wird
  - ▶ das ist die **Vorhersage**
- ▶ in Zeile 3 findet die **Korrektur** statt:
  - ▶ Vorhersage ( $\overline{bel}(x_t)$ ) wird mit der Wahrscheinlichkeit multipliziert, dass die Messung  $z_t$  in diesem Zustand wahrgenommen wird (Satz von Bayes)

$$2 \quad \overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1}$$

$$3 \quad bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t)$$



## Bayes-Filter-Algorithmus (cont.)

- ▶ Algorithmus muss mit einem *belief* zum Zeitpunkt  $t = 0$  initialisiert werden
  - ▶ falls Startzustand  $x_0$  bekannt, kann der  $bel(x_0) = 1$  und die Wahrscheinlichkeit für alle anderen Zustände gleich 0 gesetzt werden
  - ▶ falls Startzustand unbekannt, werden alle Zustände mit der gleichen Wahrscheinlichkeit initialisiert



## Bayes-Filter-Algorithmus (cont.)

- ▶ Algorithmus kann in dieser Form nur für sehr einfache Probleme implementiert werden
  - ▶ entweder muss die Integration in Zeile 2 und die Multiplikation in Zeile 3 in geschlossener Form möglich sein,
  - ▶ oder ein endlicher Zustandsraum muss gegeben sein, so dass das Integral in Zeile 2 zu einer Summe wird

$$2 \quad \overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1}$$

$$3 \quad bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t)$$



## Bayes-Filter-Algorithmus (cont.)

- ▶ für konkrete Implementierungen des Bayes-Filter-Algorithmus werden drei Wahrscheinlichkeitsverteilungen benötigt:
  - ▶ der initiale *belief*  $p(x_0)$
  - ▶ die Wahrscheinlichkeit der Messungen  $p(z_t | x_t)$
  - ▶ die Zustandsübergangswahrscheinlichkeiten  $p(x_t | u_t, x_{t-1})$



## Markov-Annahme

- ▶ Die Annahme, dass ein Zustand komplett ist, wird **Markov-Annahme** (engl. *Markov assumption*) genannt.
- ▶ Die Annahme fordert, dass vergangene und zukünftige Daten unabhängig sind, wenn man den aktuellen Zustand  $x_t$  kennt.



## Markov-Annahme (cont.)

- ▶ folgendes Beispiel verdeutlicht die Härte dieser Annahme; zur Lokalisierung mobiler Roboter werden Bayes-Filter eingesetzt:
  - ▶ dabei ist  $x_t$  die Position im Raum (Pose) des Roboters mit Bezug auf eine feste Karte
  - ▶ es gibt Effekte, die die Sensormessungen systematisch verfälschen und damit die Markov-Annahme zunichte machen:
    - ▶ Einfluss sich bewegender Personen auf die Sensormessungen
    - ▶ Ungenauigkeiten in den probabilistischen Modellen  $p(z_t | x_t)$  und  $p(x_t | u_t, x_{t-1})$
    - ▶ Rundungsfehler, wenn Näherungen für die Repräsentation des *beliefs* verwendet werden



## Markov-Annahme (cont.)

- ▶ Variablen innerhalb der Software, die mehrere Stellgrößen beeinflussen
  - ▶ meist können diese Variablen im Zustand berücksichtigt werden, was aber oft unterlassen wird, um den rechnerischen Aufwand zu verringern
- ▶ In der Praxis sind Bayes-Filter erstaunlich robust gegen die Verletzung der Markov-Annahme!

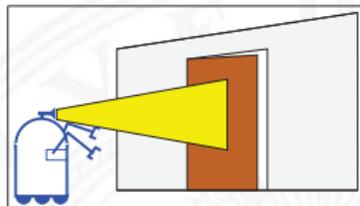


## Beispiel Bayes-Filter

Beispielszenario:

Roboter detektiert offene/geschlossene Tür (siehe [TBF05])

- ▶ Tür hat genau zwei Zustände, beschrieben durch Zustandsvariable  $\mathbf{X}$ , mit  $\Omega = \{is\_open, is\_closed\}$
- ▶ Roboter kann Zustand der Tür messtechnisch, allerdings mit Unsicherheit, erfassen



Wahrscheinlichkeiten der Detektion:

Tür offen  $p(Z_t=sense\_open \mid X_t=is\_open) = 0.6;$

$p(Z_t=sense\_closed \mid X_t=is\_open) = 0.4;$

Tür geschlossen  $p(Z_t=sense\_closed \mid X_t=is\_closed) = 0.8;$

$p(Z_t=sense\_open \mid X_t=is\_closed) = 0.2;$



## Beispiel Bayes-Filter (cont.)

- ▶ nur der Roboter kann Zustand der Tür mit Manipulator ändern

Wahrscheinlichkeiten der Aktion:

öffnen

$$p(X_t = is\_open \mid U_t = push, X_{t-1} = is\_open) = 1.0;$$

$$p(X_t = is\_closed \mid U_t = push, X_{t-1} = is\_open) = 0.0;$$

$$p(X_t = is\_open \mid U_t = push, X_{t-1} = is\_closed) = 0.8;$$

$$p(X_t = is\_closed \mid U_t = push, X_{t-1} = is\_closed) = 0.2;$$

k\_Aktion

$$p(X_t = is\_open \mid U_t = ---, X_{t-1} = is\_open) = 1.0;$$

$$p(X_t = is\_closed \mid U_t = ---, X_{t-1} = is\_open) = 0.0;$$

$$p(X_t = is\_open \mid U_t = ---, X_{t-1} = is\_closed) = 0.0;$$

$$p(X_t = is\_closed \mid U_t = ---, X_{t-1} = is\_closed) = 1.0;$$

schließen

$$p(X_t = is\_open \mid U_t = pull, X_{t-1} = is\_open) = 0.4;$$

$$p(X_t = is\_closed \mid U_t = pull, X_{t-1} = is\_open) = 0.6;$$

$$p(X_t = is\_open \mid U_t = pull, X_{t-1} = is\_closed) = 0.0;$$

$$p(X_t = is\_closed \mid U_t = pull, X_{t-1} = is\_closed) = 1.0;$$



## Beispiel Bayes-Filter (cont.)

- ▶ Zeitpunkt  $t = 0$ :
  - ▶ Initialzustand der Tür unbekannt
  - ▶ Annahme der Gleichverteilung:
    - ▶  $bel(X_0=is\_open) = 0.5$
    - ▶  $bel(X_0=is\_closed) = 0.5$





## Beispiel Bayes-Filter (cont.)

- ▶ Zeitpunkt  $t = 1$ :
  - ▶ keine Aktion ( $U_1 = ---$ )
  - ▶ Messung detektiert offene Tür ( $Z_1 = \textit{sense\_open}$ )

nach Bayes-Filter-Algorithmus

- ▶ Vorhersage ( $\overline{bel}(x_1)$ ) berechnen
- ▶ Update der Vorhersage ( $bel(x_1)$ ) durch Messung  $z_1$



## Beispiel Bayes-Filter (cont.)

**Vorhersage** berechnen nach Zeile 2 des Bayes-Algorithmus:

$$2 \quad \overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1}$$

da Zustandsraum  $\Omega$  endlich, Darstellung durch Summe:

$$\overline{bel}(x_t) = \sum_{x_{t-1}} p(x_t | u_t, x_{t-1}) bel(x_{t-1})$$

$$\begin{aligned} \overline{bel}(x_1) &= \sum_{x_0} p(x_1 | u_1, x_0) bel(x_0) \\ &= p(x_1 | U_1=--, X_0=is\_open) bel(X_0=is\_open) \\ &\quad + p(x_1 | U_1=--, X_0=is\_closed) bel(X_0=is\_closed) \end{aligned}$$

$X_1$  kann die Werte  $is\_open$  und  $is\_closed$  annehmen;

→ berechnen der Vorhersage ( $\overline{bel}$ ) für diese beiden Fälle



## Beispiel Bayes-Filter (cont.)

$$\begin{aligned} \overline{bel}(x_1) &= p(x_1 | U_1=--, X_0=is\_open) bel(X_0=is\_open) \\ &\quad + p(x_1 | U_1=--, X_0=is\_closed) bel(X_0=is\_closed) \end{aligned}$$

$$\begin{aligned} \overline{bel}(X_1=is\_open) &= p(X_1=is\_open | U_1=--, X_0=is\_open) bel(X_0=is\_open) \\ &\quad + p(X_1=is\_open | U_1=--, X_0=is\_closed) bel(X_0=is\_closed) \\ &= 1.0 \cdot 0.5 + 0.0 \cdot 0.5 \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} \overline{bel}(X_1=is\_closed) &= p(X_1=is\_closed | U_1=--, X_0=is\_open) bel(X_0=is\_open) \\ &\quad + p(X_1=is\_closed | U_1=--, X_0=is\_closed) bel(X_0=is\_closed) \\ &= 0.0 \cdot 0.5 + 1.0 \cdot 0.5 \\ &= 0.5 \end{aligned}$$

Vorhersage entspricht alter Schätzung, da keine Aktion



## Beispiel Bayes-Filter (cont.)

**Update** der Vorhersage durch Einbeziehung der Messung:

$$3 \quad \text{bel}(x_t) = \eta p(z_t|x_t) \overline{\text{bel}}(x_t) \quad (\text{Zeile 3 des Bayes-Algorithmus})$$

$$\text{bel}(x_1) = \eta p(Z_1=\text{sense\_open}|x_1) \overline{\text{bel}}(x_1)$$

$$\begin{aligned} \text{bel}(X_1=\text{is\_open}) &= \eta p(Z_1=\text{sense\_open} \mid X_1=\text{is\_open}) \overline{\text{bel}}(X_1=\text{is\_open}) \\ &= \eta \cdot 0.6 \cdot 0.5 \\ &= \eta \cdot 0.3 && = 0.75 \end{aligned}$$

$$\begin{aligned} \text{bel}(X_1=\text{is\_closed}) &= \eta p(Z_1=\text{sense\_open} \mid X_1=\text{is\_closed}) \overline{\text{bel}}(X_1=\text{is\_closed}) \\ &= \eta \cdot 0.2 \cdot 0.5 \\ &= \eta \cdot 0.1 && = 0.25 \end{aligned}$$

Berechnung des Normalisierungsfaktors  $\eta$ :

$$1 = \eta \cdot 0.3 + \eta \cdot 0.1 \quad \eta = (0.3 + 0.1)^{-1} = 2.5$$



## Beispiel Bayes-Filter (cont.)

- ▶ Zeitpunkt  $t = 2$ :
  - ▶ Aktion: *öffnen* ( $U_2 = \text{push}$ )
  - ▶ Messung detektiert offene Tür ( $Z_2 = \text{sense\_open}$ )

analog zu Periode ( $t = 1$ )  $\overline{bel}(x_2)$ , Normalisierungsfaktor und  $bel(x_2)$  berechnen:

$$\overline{bel}(X_2 = \text{is\_open}) = 1.0 \cdot 0.75 + 0.8 \cdot 0.25 = 0.95$$

$$\overline{bel}(X_2 = \text{is\_closed}) = 0.0 \cdot 0.75 + 0.2 \cdot 0.25 = 0.05$$

$$bel(X_2 = \text{is\_open}) = \eta 0.6 \cdot 0.95 = \eta 0.57 \qquad = 0.98$$

$$bel(X_2 = \text{is\_closed}) = \eta 0.2 \cdot 0.05 = \eta 0.01 \qquad = 0.02$$

$$\eta = (0.57 + 0.01)^{-1} \qquad = 1.72$$



## Mathematische Herleitung des Bayes-Filter

▶ **Beweis durch Induktion:**

zeige, dass  $\overline{bel}(x_t) = p(x_t \mid z_{1:t-1}, u_{1:t})$  aus  $p(x_{t-1} \mid z_{1:t-1}, u_{1:t-1})$  hergeleitet werden kann

▶ **Beweis-Basis:** korrekte Initialisierung von  $bel(x_0)$  zum Zeitpunkt  $t = 0$

▶ **Bedingungen:** Zustand  $x_t$  ist komplett und die Stellgrößen werden zufällig gewählt

▶ nach dem Satz von Bayes gilt nach (26) für  $bel(x_t)$ :

$$\begin{aligned} bel(x_t) = p(x_t \mid z_{1:t}, u_{1:t}) &= \frac{p(z_t \mid x_t, z_{1:t-1}, u_{1:t})p(x_t \mid z_{1:t-1}, u_{1:t})}{p(z_t \mid z_{1:t-1}, u_{1:t})} \\ &= \eta p(z_t \mid x_t, z_{1:t-1}, u_{1:t})p(x_t \mid z_{1:t-1}, u_{1:t}) \end{aligned}$$



## Mathematische Herleitung des Bayes-Filter (cont.)

$$bel(x_t) = p(x_t | z_{1:t}, u_{1:t}) = \eta p(z_t | x_t, z_{1:t-1}, u_{1:t}) p(x_t | z_{1:t-1}, u_{1:t})$$

- ▶ da der Zustand komplett ist, gilt (bedingte Unabhängigkeit)

$$p(z_t | x_t, z_{1:t-1}, u_{1:t}) = p(z_t | x_t) \quad (\text{vgl. (25)})$$

- ▶ daraus folgt:

$$bel(x_t) = p(x_t | z_{1:t}, u_{1:t}) = \eta p(z_t | x_t) p(x_t | z_{1:t-1}, u_{1:t})$$

- ▶ und mit (27)

$$bel(x_t) = \eta p(z_t | x_t) \overline{bel}(x_t),$$

was Zeile 3 des Bayes-Filter-Algorithmus entspricht



## Mathematische Herleitung des Bayes-Filter (cont.)

- ▶ Nach dem Gesetz der totalen Wahrscheinlichkeit ( $p(a) = \int p(a|c)p(c)dc$ ) gilt:

$$\begin{aligned} \overline{bel}(x_t) &= p(x_t | z_{1:t-1}, u_{1:t}) \\ &= \int p(x_t | x_{t-1}, z_{1:t-1}, u_{1:t}) p(x_{t-1} | z_{1:t-1}, u_{1:t}) dx_{t-1} \end{aligned}$$

- ▶ wenn wir  $x_{t-1}$  kennen, liefern  $z_{1:t-1}$  und  $u_{1:t-1}$  keine zusätzliche Information (Annahme: Zustand komplett)
- ▶ daher gilt die Vereinfachung

$$p(x_t | x_{t-1}, z_{1:t-1}, u_{1:t}) = p(x_t | x_{t-1}, u_t) \quad (\text{vgl. (24)})$$

$u_t$  muss verbleiben, da diese Stellgröße erst zum Zustandsübergang führt

$$\overline{bel}(x_t) = \int p(x_t | x_{t-1}, u_t) p(x_{t-1} | z_{1:t-1}, u_{1:t}) dx_{t-1}$$



## Mathematische Herleitung des Bayes-Filter (cont.)

$$\overline{bel}(x_t) = \int p(x_t | x_{t-1}, u_t) p(x_{t-1} | z_{1:t-1}, u_{1:t}) dx_{t-1}$$

- ▶ für den Übergang in den Zustand  $x_{t-1}$  ist die Steuergröße  $u_t$  irrelevant, kann als zufällig gewählt aufgefasst, und daher in den Bedingungen für  $p(x_{t-1} | z_{1:t-1}, u_{1:t})$  weggelassen werden
- ▶ es ergibt sich mit

$$\begin{aligned} \overline{bel}(x_t) &= \int p(x_t | x_{t-1}, u_t) p(x_{t-1} | z_{1:t-1}, u_{1:t-1}) dx_{t-1} \\ &= \int p(x_t | x_{t-1}, u_t) bel(x_{t-1}) dx_{t-1} \end{aligned}$$

die 2. Zeile des Bayes-Filter-Algorithmus



## Darstellung und Berechnung

- ▶ Bayes-Filter können auf unterschiedliche Weise implementiert werden
- ▶ Techniken basieren auf unterschiedlichen Annahmen über
  - ▶ Wahrscheinlichkeit der Messung
  - ▶ Wahrscheinlichkeiten der Zustandsübergänge
  - ▶ *belief*
- ▶ die Algorithmen besitzen dadurch unterschiedliche rechnerische Eigenschaften
- ▶ in den meisten Fällen müssen die *beliefs* angenähert werden
- ▶ dies hat Auswirkungen auf die Komplexität der Algorithmen
- ▶ keine der verschiedenen Techniken ist generell zu bevorzugen



## Zusammenfassung

- ▶ Die Interaktion zwischen einem Roboter und dessen Umgebung wird modelliert als gekoppeltes dynamisches System. Der Roboter setzt dazu Stellgrößen um die Umgebung zu manipulieren und nimmt die Umgebung über Sensormessungen wahr.
- ▶ Dynamik wird durch zwei wahrscheinlichkeitstheoretische Gesetzmäßigkeiten charakterisiert
  - ▶ die Wahrscheinlichkeitsverteilungen für den Zustandsübergang
  - ▶ die  $p(z_t | x_t)$  für die Messungen

Erstere beschreibt, wie der Zustand sich mit der Zeit ändert, die Zweite, wie Messungen wahrgenommen werden.



## Zusammenfassung (cont.)

- ▶ der **belief** ist die A-Posteriori-Wahrscheinlichkeit des Zustands; gegeben: alle bisherigen Messungen und Stellgrößen
  - ▶ *belief* wird mittels *Bayes-Filter* berechnet
  - ▶ es ist ein rekursiver Algorithmus
  - ▶ der *belief* zum Zeitpunkt  $t$  wird aus dem *belief* zum Zeitpunkt  $t-1$  berechnet
- ▶ der Bayes-Filter basiert auf *Markov-Annahme*
  - ▶ der Zustand ist eine komplette Zusammenfassung der Vergangenheit
  - ▶ Annahme trifft in der Praxis meist nicht zu
- ▶ Bayes-Filter sollten vor ihrem Einsatz in speziellen Anwendungen anhand verschiedener Kriterien wie Genauigkeit, Effizienz und Einfachheit evaluiert werden.



# Lokalisierung

- ▶ Fähigkeit eines Roboters, seine Position in Bezug zu einem globalen Koordinatensystem zu bestimmen
- ▶ lokale Lokalisierung
  - ▶ Startposition ist bekannt
  - ▶ Positionsbestimmung nach Aktion
  - ▶ Bezugspunkt ist Startposition (relative Lokalisierung)
  - ▶ Positionstracking
- ▶ globale Lokalisierung
  - ▶ Startposition unbekannt
  - ▶ absolute Lokalisierung
  - ▶ Kidnapped-Robot, Wake-up-Robot



## Beispiele für Bayes-Filter in der Lokalisierung

- ▶ Kalman-Filter
  - ▶ Fehlerverteilung der Messfehler durch Gaussverteilung modelliert
  - ▶ für lokale Lokalisierungsprobleme geeignet
  - ▶ Veränderung des Zustandes linear und a-priori bekannt
- ▶ Histogramm-Filter
  - ▶ Diskretisierung des Zustandsraumes
  - ▶ ermöglicht multimodale Verteilungen
- ▶ Partikelfilter
  - ▶ von gitterbasierter Lokalisierung abgeleitet
  - ▶ keine Annahme der Normalverteilung des geschätzten Wertes
  - ▶ sowohl für lokale wie auch für globale Lokalisierung geeignet
  - ▶ Schätzung wird als Wahrscheinlichkeitsverteilung modelliert
  - ▶ Verteilung wird an ausgewählten Stellen, den sog. Partikeln diskretisiert und zwischen diesen wird interpoliert



## Effizienz

- ▶ einige Näherungen bedingen polynomiale Laufzeiten in Abhängigkeit von der Dimension des Zustands (z.B. Kalman-Filter)
- ▶ einige haben exponentielle Laufzeiten (z. B. gridbasierte Techniken)
- ▶ partikel-basierte Verfahren haben eine Laufzeit, die von der gewünschten Genauigkeit abhängt



## Genauigkeit

- ▶ einige Näherungen eignen sich besser, um eine Reihe von Wahrscheinlichkeitsverteilungen anzunähern
  - ▶ beispielsweise eignen sich Normalverteilungen für uni-modale Wahrscheinlichkeitsverteilungen
  - ▶ Histogramme können multimodale Verteilungen besser annähern, aber auf Kosten der Genauigkeit
  - ▶ Partikel-Techniken können eine Vielzahl von Verteilungen annähern, wodurch allerdings die Anzahl der Partikel sehr groß werden kann



## Markov-Lokalisierung

- ▶ Wahrscheinlichkeitstheoretische Lokalisierungsverfahren sind Varianten des Bayes-Filters
- ▶ direkte Anwendung des Bayes-Filters wird **Markov-Lokalisierung** genannt
- ▶ Markov-Lokalisierungen benötigen zusätzlich eine Karte  $m$  der Umgebung
- ▶ Karte  $m$  für das Messmodell erforderlich (teilweise auch im Bewegungsmodell berücksichtigt)
- ▶ Markov-Lokalisierung wird verwendet für
  - ▶ Positions-Tracking (lokale Lokalisierung)
  - ▶ globale Lokalisierung
  - ▶ das Kidnapped-Robot-Problem



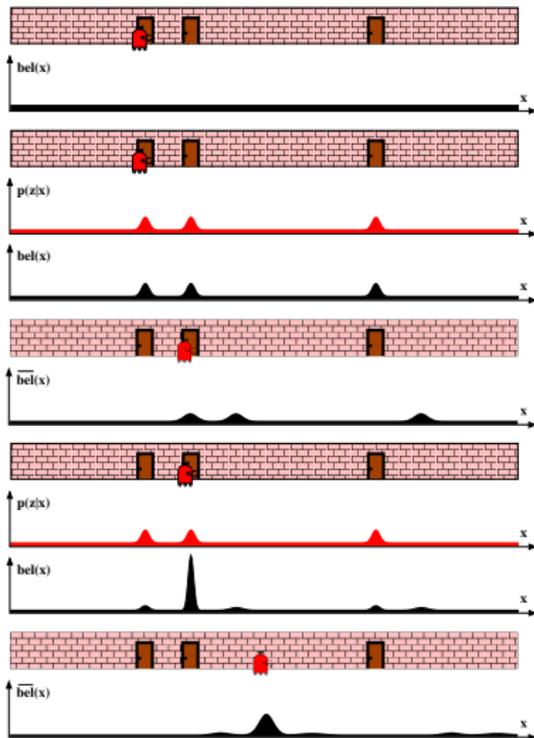
## Markov-Lokalisierung (cont.)

Markov\_Lokalisierung( $bel(x_{t-1}), u_t, z_t, m$ ):

1. **for all**  $x_t$  **do**
2.      $\overline{bel}(x_t) = \int p(x_t | u_t, x_{t-1}, m) bel(x_{t-1}) dx_{t-1}$
3.      $bel(x_t) = \eta p(z_t | x_t, m) \overline{bel}(x_t)$
4. **endfor**
5. **return**  $bel(x_t)$



## Markov-Lokalisierung (cont.)



- falls Anfangsposition und -pose unbekannt, wird für  $bel(x_0)$  eine Gleichverteilung angenommen und

$$bel(x_0) = \frac{1}{|X|}$$

gesetzt, mit  $|X|$  entspricht der Anzahl der gültigen Posen in der Karte

- falls Anfangspose zwar bekannt, aber nur als Approximation, wird  $bel(x_0)$  mit einer Gaussverteilung initialisiert:

$$bel(x_0) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(x - \bar{x}_0)\Sigma^{-1}(x - \bar{x}_0)^T \right\}$$

- grundsätzlich ist die Markov-Lokalisierung unempfindlich gegenüber der Repräsentation der zugrunde liegenden Verteilungen

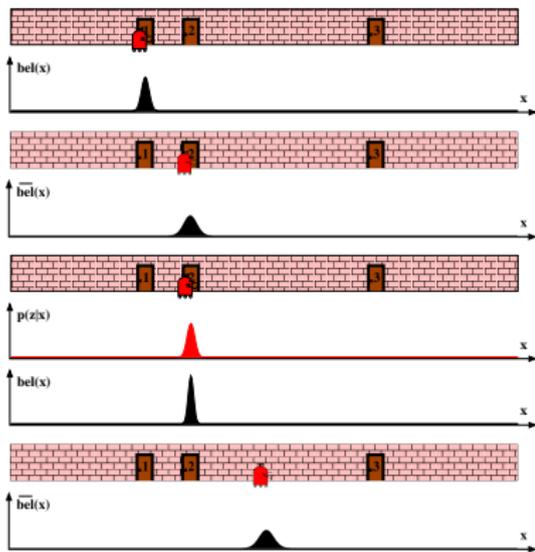


## Extended Kalman Filter Lokalisierung

- ▶ **extended Kalman filter localisation** (oder EKF-Lokalisierung) ist eine spezielle Ausprägung der Markov-Lokalisierung
- ▶ typische Anwendung: Tracking
- ▶ die EKF-Lokalisierung modelliert den *belief*  $bel(x_t)$  durch unimodale Verteilung mit Mittelwert  $\mu_t$  und Kovarianzmatrix  $\Sigma_t$
- ▶ die Karte ist dabei eine Sammlung von Merkmalen
- ▶ der Roboter misst die Entfernung und Richtung der Merkmale (Feature) in seiner Nähe
  - ▶ erstellt Merkmalsvektor  $z_t = (z_t^1, z_t^2, \dots, z_t^m)$  mit  $m = \#$ Merkm. zum Zeitpunkt  $t$ ; Merkmal  $z_t^i$  bedeutet: Merkmal  $i$  zum Zeitpunkt  $t$
  - ▶ berechnet Übereinstimmungsvektor  $c_t = (c_t^1, c_t^2, \dots, c_t^m)$ ;  $c_t^i = a$ , mit  $a \in \{1, \dots, n\}$  und  $n$  Anzahl der Merkmale der Karte, bedeutet: Merkmal  $i$  des Merkmalsvektors  $z$  zum Zeitpunkt  $t$  stimmt mit Merkmal  $a$  der Karte überein



# Extended Kalman Filter Lokalisierung (cont.)



Beispiel EKF:

► Annahmen:

- Startpose hinreichend genau bekannt (durch Gaussverteilung abgebildet)
- Korrespondenzen (jeweilige Übereinstimmungsvektoren  $c_t$  nach der Messung in  $x_t$ ) bekannt
- Wahrscheinlichkeit der Messung:  $p(z_t \mid x_t, m, c_{t-1})$
- Zustandsübergangswahrscheinlichkeit:  $p(x_t \mid x_{t-1}, m, c_{t-1})$



# Grid-Lokalisierung

**Grid-Lokalisierung** approximiert den *belief* durch einen **Histogram-Filter** über die Grid-Dekomposition des Zustandsraumes

- ▶ Diskretisierung des Zustandsraumes durch Grid-Zellen  $x$
- ▶ erlaubt multimodale Verteilungen
- ▶ dieser diskrete Bayes-Filter verwaltet eine Menge diskreter Wahrscheinlichkeitswerte:

$$bel(x_t) = \{p_{k,t}\}$$

wobei jedes  $p_{k,t}$  zu einer Grid-Zelle  $x_k$  gehört

- ▶ die Vereinigung der Grid-Zellen zum Zeitpunkt  $t$  bildet den Zustandsraum  $X_t$ :

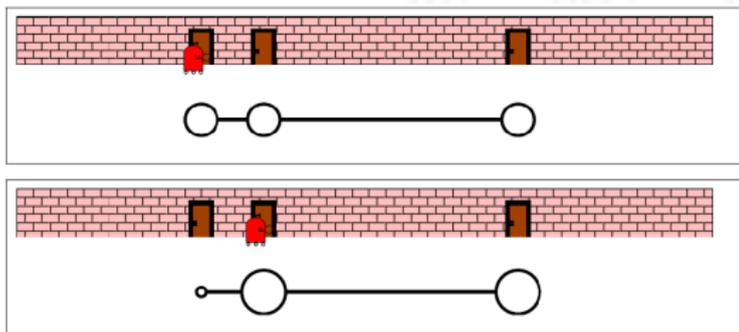
$$domain(X_t) = \{x_{1,t} \cup x_{2,t} \cup \dots \cup x_{K,t}\}$$

## Grid-Lokalisierung (cont.)

- ▶ zwei Varianten der Grid-Zerlegung üblich:

### (1) topologische Grid-Zerlegung

- ▶ Zelle entspricht signifikanten Ort/Merkmal in der Karte (Flurbeispiel: Tür, Einmündung ...)
- ▶ üblicherweise sehr grobes Grid
- ▶ Grid abhängig von Gegebenheiten

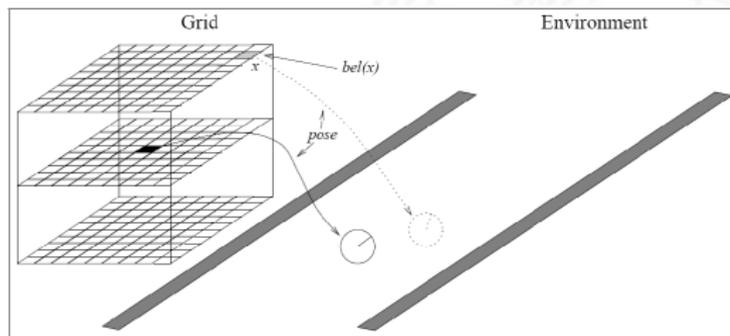


## Grid-Lokalisierung (cont.)

- ▶ zwei Varianten der Grid-Zerlegung üblich:

### (2) metrische Grid-Zerlegung

- ▶ gleiche Zellengröße
- ▶ typische Werte der Zellengröße in der Praxis sind bis zu 15cm Tiefenauflösung bei bis zu 5° Winkelauflösung
- ▶ höhere Auflösung gegenüber topologischem Grid zulasten des Aufwandes





## Grid-Lokalisierung (cont.)

**Grid\_Lokalisierung**( $\{p_{k,t-1}, u_t, z_t, m\}$ ):

1. **for all**  $k$  **do**

2.

$$\bar{p}_{k,t} = \sum_i [p_{i,t-1} \cdot \text{Bewegungsmodell}(\text{mean}(x_k), u_t, \text{mean}(x_i))]$$

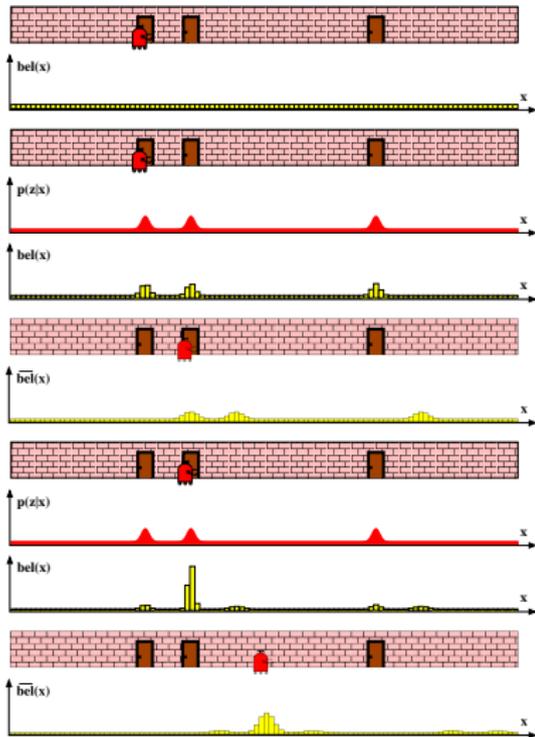
3.  $p_{k,t} = \eta \cdot \text{Messmodell}(z_t, \text{mean}(x_k), m) \cdot \bar{p}_{k,t}$

4. **endfor**

5. **return**  $p_{k,t}$

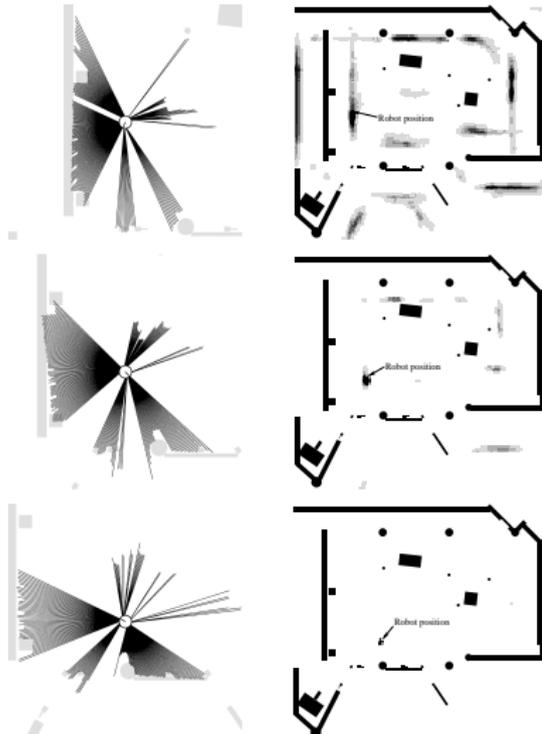
die Funktion *mean* berechnet den physikalischen Schwerpunkt einer Zelle  $x_i$

## Grid-Lokalisierung (cont.)



- ▶ feingliedrige metrische Zerlegung
- ▶ Annahme der Gleichverteilung für den Startzustand  $x_0$
- ▶ belief  $bel(x_t)$  dargestellt als Histogramm der Zelle
- ▶ zugrundeliegende Verteilungen können mehrdimensional sein

## Grid-Lokalisierung (cont.)



Beispiel: globale Lokalisierung

- ▶ Winkelauflösung  $5^\circ$ ,  
Tiefenauflösung 15 cm
- ▶ Annahme der Gleichverteilung für  
den Startzustand  $x_0$
- ▶ Messwerte außerhalb der  
Reichweite wurden fortgelassen
- ▶ graue Bereiche sind Orte erhöhter  
Aufenthaltswahrscheinlichkeit
- ▶ belief dreidimensional  $(x, y, \theta)$ ; zur  
Vereinfachung jedoch nur in den  
 $(x, y)$ -Raum projiziert



# Monte-Carlo-Lokalisierung

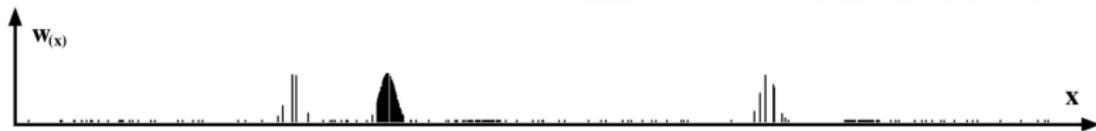
- ▶ **Monte-Carlo-Lokalisierung** (MCL) approximiert den *belief*  $bel(x_t)$  durch eine Menge  $S_t$  von Partikeln
- ▶ Partikelmenge (Samples):  $S_t = \{s^{(i)} | s^{(i)} \in X_t\}$  mit  $i = 1 \dots N$  und  $X_t$  Zustandsraum in  $t$ 
  - ▶ diskrete Näherung einer Wahrscheinlichkeits-Verteilung
  - ▶ Annäherung nahezu beliebiger Verteilungsfunktionen möglich
  - ▶ es können multimodale und unimodale Verteilungen approximiert und übergangslos zwischen ihnen gewechselt werden
- ▶ Wichtung der Samples  $\bar{s}^{(i)} = (s^{(i)}, w^{(i)})$  mittels Messmodell





## Monte-Carlo-Lokalisierung (cont.)

- ▶ eine höhere Anzahl von Partikeln erhöht die Genauigkeit
- ▶ Partikelmengengröße kann zusätzlich adaptiv verändert werden
  - ▶ wenn z. B. hohe Übereinstimmung mit Sensorwerten, d. h. Wichtungsfaktoren hoch, kann Samplemenge verkleinert werden
- ▶ Berechnungsverfahren strukturell ähnlich Markov-Lokalisierung, jedoch kein Bayes-Filter
  - ▶ zusätzlich **Resampling**; Wahl der neuen Samplemenge  $S_t$ 
    - ▶ aus Elementen der alten Samplemenge  $S_{t-1}$
    - ▶ ggf. mit Generierung neuer Sample





## Monte-Carlo-Lokalisierung (cont.)

**MC\_Lokalisierung**( $S_{t-1}, u_t, z_t, m$ ):

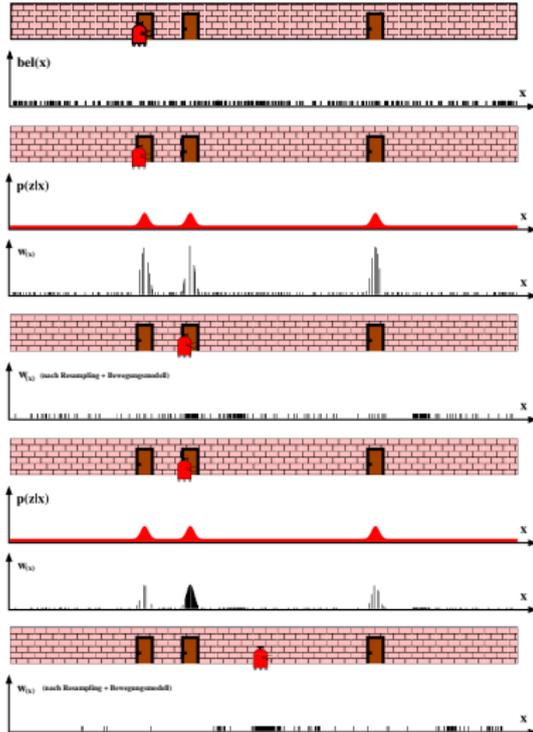
1.  $\bar{S}_t = S_t = \emptyset$
2. **for**  $k = 1$  **to**  $N$  **do**
3.  $x_t^{(k)} = \text{Bewegungsmodell}(u_t, x_{t-1}^{(k)}, m)$
4.  $w_t^{(k)} = \text{Messmodell}(z_t, x_t^{(k)}, m)$
5.  $\bar{S}_t = \bar{S}_t \cup \{(x_t^{(k)}, w_t^{(k)})\}$
6. **endfor**
7. **for**  $k = 1$  **to**  $N$  **do**
8.  $\text{wählePartikel}(i, \bar{S}_t)$
9.  $S_t = S_t \cup \{s_t^{(i)}\}$
10. **endfor**
11. **return**  $S_t$

*% N = # Partikel*

die Funktion *wählePartikel* wählt z. B. quasi zufällig, aber mit Bevorzugung hoher Gewichte, einen Partikel mit Zurücklegen aus  $\bar{S}_t$  aus; bei Wahl eines bereits aktivierten Partikels, wird z. B. der nächstgelegene linke oder rechte Partikel aktiviert



# Monte-Carlo-Lokalisierung (cont.)





[TBF05] Sebastian Thrun, Wolfram Burgard, Dieter Fox,  
Ronald C. Arkin (Hrsg.):  
*Probabilistic Robotics*.  
MIT Press; Cambridge, Massachusetts, 2005  
(Intelligent Robotics and Autonomous Agents). –  
Kapitel 2