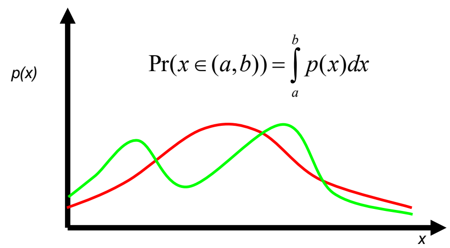
Uncertainty and Error

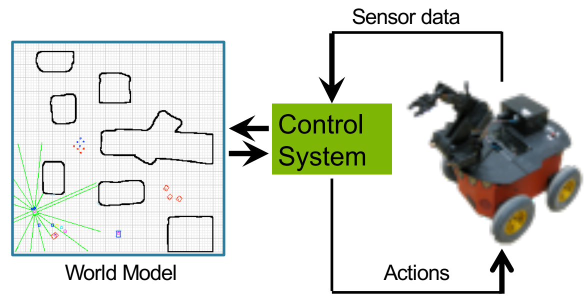
$$\Pr(x \in (a,b)) = \int_a^b p(x)dx$$



Many slides adapted from slides © R. Siegwart, Steve Seitz, J. Tim Oates

1

High Level View



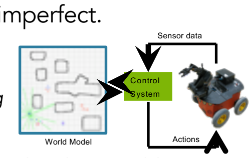
2

2

Uncertainty in Robotics

- Fundamentally, models are imperfect.
  - Sensors aren't perfect
  - Actuation isn't either
  - But you have to do *something*
- Probability as uncertainty
  - Probability theory can be applied to these problems
- Key idea: explicit representation of uncertainty using the calculus of probability theory

Perception = state estimation
Action = utility optimization



3

3

Error and Uncertainty

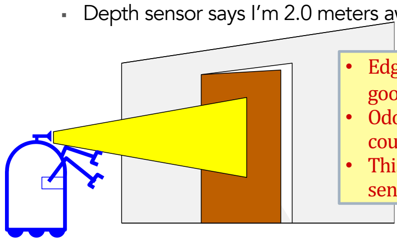
- Sensing is always related to uncertainty.
- What are the sources of uncertainties?
  - Blown-out camera; iffy rangefinder; skidding wheel; background noise; poor speech model; what else?
- How can uncertainties be represented / quantified
  - Deterministic vs. random error
- How do they propagate?
  - Uncertainty of a function of uncertain values?
  - How do uncertainties combine if different sensor reading are fused?

4

4

Example: State Estimation

- Is the door open?
  - Camera + edge detection says the door is not at right angles
  - Odometry says I'm 2.0 meters away from door frame
  - Depth sensor says I'm 2.0 meters away from door



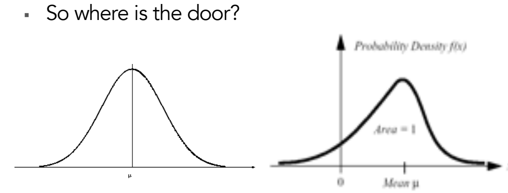
- Edge detection pretty good indoors?
- Odometry very noisy; could be off by 20cm.
- This specific depth sensor is very good

5

5

Distributions

- How can a reading be wrong?
  - Poor surface for your distance sensor
  - You may be using an imprecise ranging method
  - Someone walked in front of it
  - So where is the door?

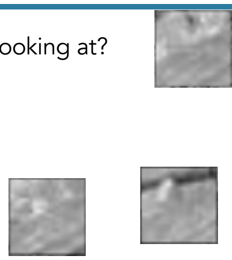


6

6

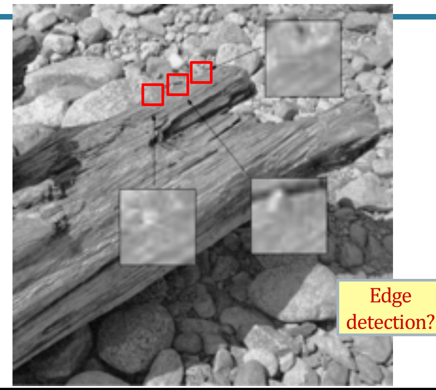
Vision

- What are we looking at?



7

Vision



8

Rational decisions & uncertainty

- Making rational decisions under uncertainty
  - *Probability*
    - the precise representation of knowledge and uncertainty
  - *Probability theory*
    - How to optimally update your knowledge based on new information
  - *Decision theory: probability theory + utility theory*
    - How to use this information to achieve maximum expected utility

9

Decision-Making + Utility

- Decision theory: probability theory + utility theory
  - How to use this information to achieve maximum *expected* utility ("Goodness")
- Consider a bus schedule. What's the utility function?
  - A schedule says the bus comes at 8:05.

    Situation A: You have a class at 8:30.

    Situation B: You have a class at 8:30, and it's cold and raining.

    Situation C: You have a final exam at 8:30, it's cold and raining.

10

Discrete Random Variables

- X denotes a random variable.
- X can take countable number of values in $\{x_1, x_2, \ldots, x_n\}$
- $P(X=x_i)$ or $P(x_i)$ or $Pr(x_i)$ is the probability that the random variable X takes on value x_i.
- $P(\cdot)$ is called its probability mass function.
- E.g.

$$P(RoomType) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$$

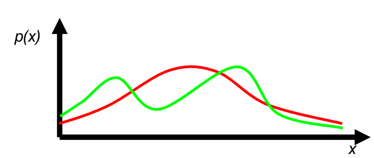
11

Continuous Random Variables

- X takes on values in the continuum.
- $p(X=x)$, or $p(x)$, is a probability density function.

$$\Pr(x \in (a,b)) = \int_a^b p(x)dx$$

- E.g.



12

Axioms of Probability

- Pr(*A*) denotes probability that proposition *A* is true.
- Axioms (Kolmogorov): $0 \le P(A) \le 1$
  $P(True) = 1 \quad P(False) = 0$
  $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
- Corollaries:
  - A random variable must sum to one: $\sum_{i=1}^{n} P(D = d_i) = 1$
  - The joint probability of a set of variables must also sum to one
- If A and B are mutually exclusive: $P(A \vee B) = P(A) + P(B)$

13

13

Conditionality

- $P(B|A)$
  - Probability of event B given Event A

    - aka -
  - Pretend A has already happened, or we know how likely it is to happen. Now what is the chance of event B? (For all possible values of A?)
- $P(B \mid A)$ is the "Conditional Probability" of B given A

14

14

Rules of Probability

- Conditional probability

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}, \qquad P(B) > 0$$

- Corollary: **Bayes Law**

$$P(B|A)\,P(A) = P(A \text{ and } B) = P(A|B)\,P(B)$$

$$\Rightarrow \quad P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} = \frac{\text{likelihood} \bullet \text{prior}}{\text{evidence}}$$

Probability of an event based on a *prior:* Conditions that may relate to that event

15

15

Bayes!

- Probability of an **event** based on **conditions** that may relate to that event

$$P(x,y) = P(x \mid y)P(y) = P(y \mid x)P(x)$$

$$\Rightarrow$$

$$P(x|y) = \frac{P(y \mid x)\,P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

16

16

Independence

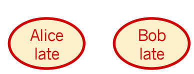
- Two variables X, Y are independent when the probability of X is not related to the probability Y:

$$P(x|y) = P(x)$$

and

$$P(x \text{ and } y) = P(x) \cdot P(y)$$

for all values of X and Y



- Is Alice late to work? Is Bob late to work?

17

17

Conditional Dependence

- Two variables X, Y are *conditionally dependent* when P(X) and P(Y) each depend on a third factor, P(Z):

$$P(x,y|z) = P(x \mid z)P(y \mid z)$$

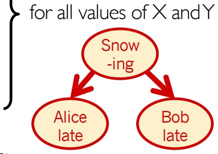
$$\Leftrightarrow$$

$$P(x|z) = P(x \mid z, y)$$

and

$$P(y|z) = P(y \mid z, x)$$

for all values of X and Y



- Alice late / Bob late / Snowing

18

18

Bayes + Background Knowledge

- Probability of an event based on conditions that may relate to that event
- Example: Does Alice have cancer?
  - Alice is 65
  - ***If*** *cancer is related to age*, we can use that knowledge to improve accuracy of our assessment using Bayes

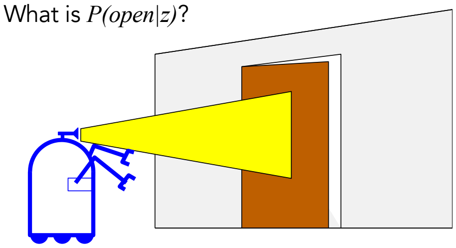
$$P(x \mid y,z) = \frac{P(y \mid x,z)\ P(x \mid z)}{P(y \mid z)}$$

19

19

State Estimation

- Suppose a robot obtains **measurement** z
  - z = vision + edge detection
- What is $P(open|z)$?



21

21

Casual (Observed) Priors

- $P(open|z)$ is **diagnostic**.
- $P(z|open)$ is **causal**.
- Often **causal** knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge:

count frequencies!

$$P(open \mid z) = \frac{P(z \mid open) P(open)}{P(z)}$$

22

22

Example

- $P(z|open) = 0.6$ $\quad P(z|\neg open) = 0.3$
- $P(open) = P(\neg open) = 0.5$

$$P(open \mid z) = \frac{P(z \mid open) P(open)}{P(z \mid open) p(open) + P(z \mid \neg open) p(\neg open)}$$

$$P(open \mid z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{2}{3} = 0.67$$

This z gives higher probability that the door is open.

23

23

Combining Evidence

- Suppose our robot obtains another observation z_2.
- How can we integrate this new information?
- More generally, how can we estimate $P(x| z_1 \ldots z_n)$?

24

24

Second Measurement

- $P(z_2|open) = 0.5$ $\quad P(z_2|\neg open) = 0.6$
- $P(open|z_1) = 2/3$

$$P(open \mid z_2, z_1) = \frac{P(z_2 \mid open)\ P(open \mid z_1)}{P(z_2 \mid open)\ P(open \mid z_1) + P(z_2 \mid \neg open)\ P(\neg open \mid z_1)}$$

$$= \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2} \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{1}{3}} = \frac{5}{8} = 0.625$$

z_2 gives higher probability that the door is open.

25

25

Recursive Bayesian Updating

$$P(x \mid z_1,\ldots,z_n) = \frac{P(z_n \mid x, z_1,\ldots,z_{n-1})\, P(x \mid z_1,\ldots,z_{n-1})}{P(z_n \mid z_1,\ldots,z_{n-1})}$$

Markov assumption: z_n is independent of $z_1,\ldots,z_{n-1}$ if we know x

$$P(x \mid z_1,\ldots,z_n) = \frac{P(z_n \mid x)\, P(x \mid z_1,\ldots,z_{n-1})}{P(z_n \mid z_1,\ldots,z_{n-1})}$$

$$= \eta\, P(z_n \mid x)\, P(x \mid z_1,\ldots,z_{n-1})$$

$$= \eta_{1\ldots n} \prod_{i=1\ldots n} P(z_i \mid x)\, P(x)$$

$P(B|A)$: probability of B given A

26

26

Statistics Review

- Expected value of a real-valued random variable X with density f(x):
  - E[X] = ∫x f(x)
- Expected value of a discrete-valued random variable X with distribution P(x):
  - E[X] = Σx P(x)
  - Suppose X corresponds to outcome of die roll:
    - E[X] = 1 * 1/6 + 2 * 1/6 + 3 * 1/6 + 4 * 1/6 + 5 * 1/6 + 6 * 1/6
    - E[X] = 1/6 * (1 + 2 + 3 + 4 + 5 + 6) = 3.5
- If random variables X_1 and X_2 are independent:
  - $E[X_1 * X_2] = E[X_1] * E[X_2]$

27

27

Statistics Review

- Variance: how far a set of numbers is spread out.
  - E[(x - μ)2] = ∫ x2 f(x) - μ2
  - recall μ is the mean value
- If the variables are correlated, then we have covariance
- Covariance
  - Given two random variables, X1 and X2
  - E[(X1 - μX1) (X2 - μX2)]
  - What happens in the following case?
    - When X1 is above its mean, X2 tends to be below its mean
    - When X1 is above its mean, X2 tends to be way above its mean

28

28

Error and Accuracy

- Error: Difference between sensor output and true value

$$error = m\text{-}v \quad \begin{cases} m = \text{measured value} \\ v = \text{true value} \end{cases}$$

Ex: 1.2 meters

- Accuracy: a unitless measure

error

$$\left(accuracy = 1 - \frac{m - v}{v} \right)$$

Adapted from © R. Siegwart, ETH Zürich – ASL

29

29

Precision (But Not as in Recall)

- Precision: Reproducibility of sensor results
- A distribution of error can be characterized by:
  - Mean error: μ
  - Standard deviation: **σ**
  - How similar are two outputs from the same test?
    - Same sensor, same environment …

$$precision = \frac{range}{\sigma}$$

- Has other meanings in actuation and cognition

30

30

Statistical Representation of E

- Error: the difference between measured and true value
- How can we treat sensing as estimation?
- X: random variable representing actual value
  - E.g., "distance = 4 meters"
- E[X]: estimate of the true value
- Given n sensor readings (ρ1, ρ2, …, ρn)
- E[X] = g(ρ1, ρ2, …, ρn)

31

31

Representation of Uncertainty

- Specific errors are usually unknown, but…
- Errors exist on a spectrum:
  - Deterministic ⟷ Non-deterministic (random)
  - Some errors are consistent for some circumstances, and can be characterized. These are more deterministic.
  - A *probability density function* gives a probability density f(x) for any x in X.

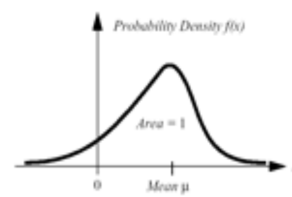
32

32

Representing Uncertainty

- Sensing as estimation problem:

  true (unknown) value = X
  estimate of value = $E[X]$
- Given n measurements with values : σ[1-n]



33

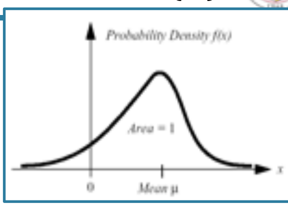
33

Uncertainty Representation (2)

Area under curve = 1:

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

sum of all possible probability values.



Mean:

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

…if we measure X infinite times and average the values we see.

Variance:

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

The "width" of possible values X might take.

34

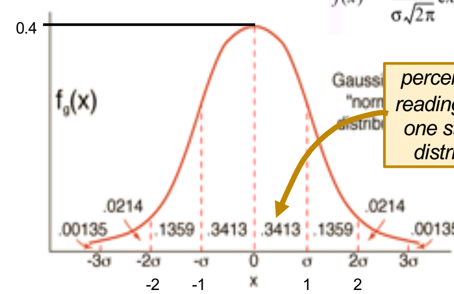
34

Gaussian Distribution

$\mu = 0$ and $\sigma = 1$

formula for Gaussian

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



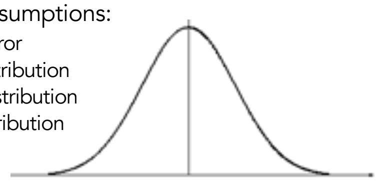
percentage of readings within one standard distribution

35

35

Error Distributions

- Random errors: behavior of sensors modeled by some probability distribution
- Causes and behavior of error usually unknown
  - So what do we do?
- Simplifying assumptions:
  - Zero-mean error
  - Unimodal distribution
  - Symmetric distribution
  - Gaussian distribution



36

36

Simplifying Assumptions

- Important to remember assumptions are wrong!
- Examples:
  - Sonar (ultrasonic) sensor more likely to overestimate distance in real environment
  - Is therefore not symmetric
    - Might be better modeled by two modes:
      - Mode for the case that the signal returns directly
      - Mode for the case that the signals returns after reflections
  - Stereo vision system might not correlate images
    - Results that make no sense at all

37

37

Error Propagation

- How do we combine a series of uncertain measurements?
  - (Basically the usual case for sensing)
- **Propagation** of uncertainty (or propagation of error)
- **Fuse** a sequence of readings into a single value

38

38

Error Propagation Law

- The effect of variables' uncertainty on the uncertainty of a function that depends on them.

  Absolute error Δx
- Error on some quantity, Δx, is given as

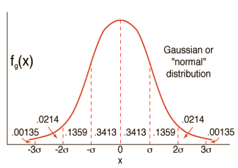
  Standard deviation: the positive square root of variance, σ2
- With a probability distribution, can find confidence *limits*
  - How sure are we of our estimate?

39

39

Confidence Limits

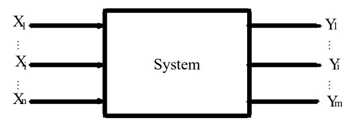
- **With a probability distribution**, can find confidence *limits*
- Example:
  - The 68% confidence limits for a one-dimensional variable in a normal distribution are ± one std. dev. from the value
  - Approximately a 68% probability that the true value lies in the region $x \pm \sigma$



40

40

The Error Propagation Law



- Error propagation in a multi-input multi-output system with n inputs and m outputs.

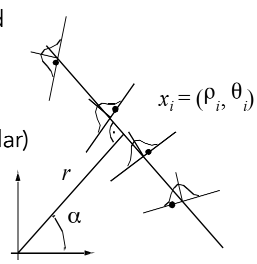
$$Y_j = f_j(X_1 \dots X_n)$$

41

41

Error Propagation Law

- Try extracting a line based on point measurements with uncertainties.
- The model parameters ri (length of the perpendicular) and qi (its angle to the abscissa) describe a line uniquely.
- The question:
  - What is the uncertainty of the extracted line knowing the uncertainties of the measurement points that contribute to it?

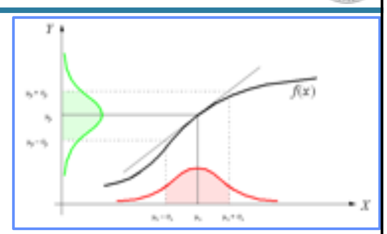


42

42

The Error Propagation Law

- One-dimensional case of an error propagation problem
- The output covariance matrix C_Y is given by the error propagation law: $C_Y = F_X C_X F_X^T$
- where
  - C_X: covariance matrix representing the input uncertainties
  - C_Y: covariance matrix representing the propagated uncertainties for the outputs.
  - F_X: is the Jacobian matrix defined as:



$$F_X = \nabla f = \left[\nabla_X \cdot f(X)^T\right]^T = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial X_1} & \cdots & \frac{\partial}{\partial X_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial X_1} & \cdots & \frac{\partial f_1}{\partial X_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial X_1} & \cdots & \frac{\partial f_m}{\partial X_n} \end{bmatrix}$$

43

43