

Lecture 3: Probability Theory Refresher

[Resources: 6.825 Techniques in Artificial Intelligence, MIT OCW
16-811 Math Fundamentals for Robotics, CMU]

CSCI 545 Introduction to Robotics
Instructor: Stefanos Nikolaidis

Motivation

- Why do we need probability?
 - To understand how likely an outcome is!

Motivation



Motivation

Is it deterministic or non-deterministic?



Motivation

Is it deterministic or non-deterministic?



From a Bayesian perspective, it doesn't matter!

Probability Theory

- It is a logic; a language that speaks about likelihood of events.
- We can start with a universe of atomic events: things that could happen or ways that the world could be
- Then, a probability distribution is a function that maps events into a range between zero and one: $P: \text{events} \rightarrow [0, 1]$

Probability Theory Axioms

- $P(\text{True}) = 1 = P(U)$

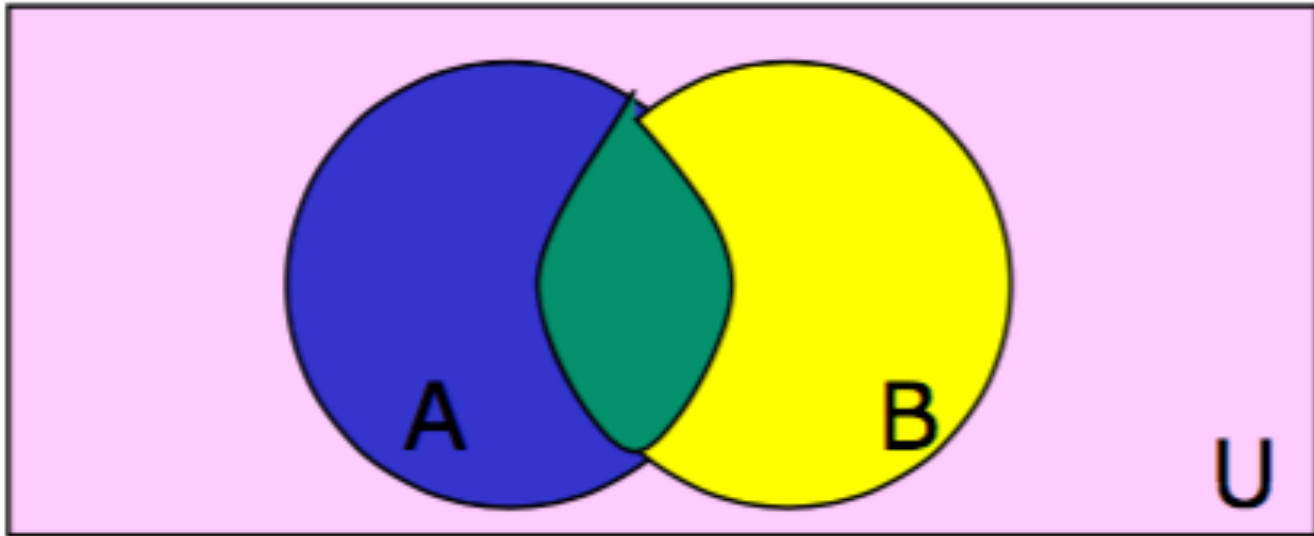
The probability of True is one. It is the probability of the universe, the probability that something in this realm of discussion that we have available to use is one.

- $P(\text{False}) = 0 = P(\text{null})$

The probability of False is 0. In terms of atomic events, False is the empty set.

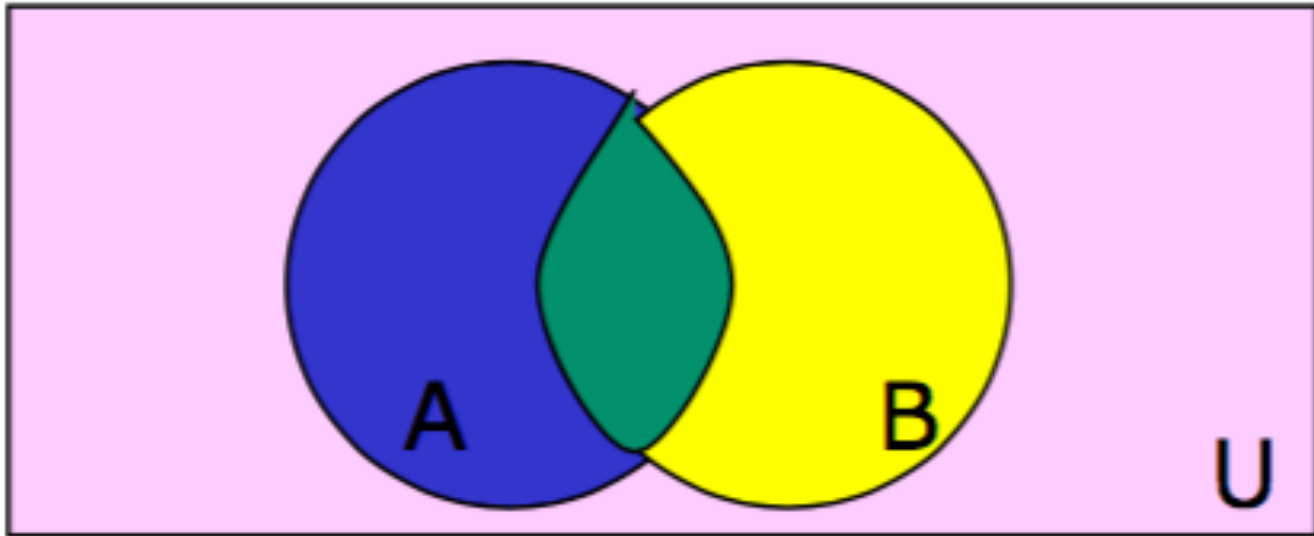
Probability Theory Axioms

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Probability Theory Axioms

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Random Variables

- A random variable: it can be considered as a mapping from a discrete domain to a range from 0 to 1.
- It sums to 1 over the domain
- Example RV: Raining
 - Domain: True / False
 - Raining (True) = 0.2

Random Variables

- $\text{Raining}(\text{True}) = 0.2$
- $P(\text{Raining} = \text{True}) = 0.2$
- $P(\text{Raining} = \text{False}) = ?$

Random Variables

- $P(\text{Raining} = \text{True}) = 0.2$
- $P(\text{Raining} = \text{True}) = 0.2$
- $P(\text{Raining} = \text{False}) = \mathbf{0.8}$

Multiple Random Variables

	Light Off	Light On
Battery Low	0.04	0.06
Battery High	0.01	0.89



Multiple Random Variables

	Light Off	Light On
Battery Low	0.04	0.06
Battery High	0.01	0.89

$$P(\text{Battery} = \text{low}) = 0.1$$

$$P(\text{Battery} = \text{high}) = 0.9$$

Multiple Random Variables

	Light Off	Light On
Battery Low	0.04	0.06
Battery High	0.01	0.89

$P(\text{Light} = \text{off}) = ?$

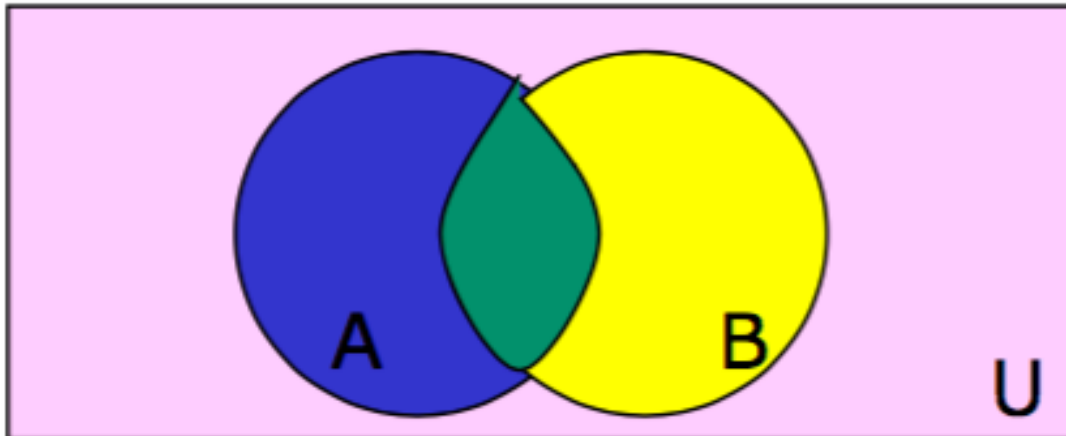
Multiple Random Variables

	Light Off	Light On
Battery Low	0.04	0.06
Battery High	0.01	0.89

$$P(\text{Light} = \text{off}) = \mathbf{0.05}$$

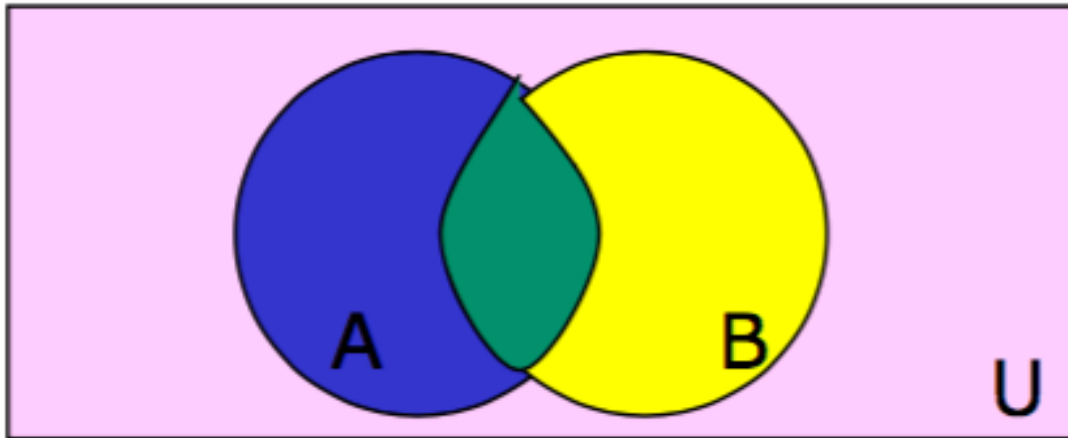
Conditional Probability

- $P(A | B)$?



Conditional Probability

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



We only consider the part of the world, where B is True.

Conditional Probability

	Light Off	Light On
Battery Low	0.04	0.06
Battery High	0.01	0.89

$$P(\text{Battery-Low}|\text{Light-Off}) =$$

Conditional Probability

	Light Off	Light On
Battery Low	0.04	0.06
Battery High	0.01	0.89

$$P(\text{Battery-Low}|\text{Light-Off}) = \frac{0.04}{0.05} = 0.8$$

Posterior probability

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

$$P(\text{State}|\text{Observation}) = \frac{P(\text{Observation}|\text{State})P(\text{State})}{P(\text{Observation})}$$

Posterior probability

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

$$P(\text{State}|\text{Observation}) = \frac{P(\text{Observation}|\text{State})P(\text{State})}{P(\text{Observation})}$$

Independent Events

- Knowing that B is true doesn't give us any more information about A.

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

Independent Events

- Are temperature and time of the day independent?
- Are the state of the battery and the whether the motor is damaged or not independent?

Conditional Independence

- A and B are conditionally independent given C, iff the probability of A given B and C is equal to the probability of A given C:

$$P(A|B, C) = P(A|C)$$

Conditional Independence

- A and B are conditionally independent given C, iff the probability of A given B and C is equal to the probability of A given C:

$$P(A|B, C) = P(A|C)$$

Conditional Independence: Example

- Random Variables:
 - Low Battery (B)
 - Light Off (L)
 - No Motion (N)
- What is the relationship of low battery, light off and no motion?

Conditional Independence: Example

- Random Variables:
 - Low Battery (B)
 - Light Off (L)
 - No Motion (N)
- What is the relationship of low battery, light off and no motion?
- Does seeing the robot with a light off tell us whether it is going to move?

Conditional Independence: Example

- Random Variables:
 - Low Battery (B)
 - Light Off (L)
 - No Motion (N)
- What is the relationship of low battery, light off and no motion?
- Does seeing the robot with a light off tell us whether it is going to move?
- What if we know that the battery is low? Does seeing a light off give us some additional information?

Incorporating Multiple Sources of Information

- Random Variables:
 - Low Battery (B)
 - Light Off (L)
 - No Motion (N)

$$P(B|L, N) = \frac{P(L, N|B)P(B)}{P(L, N)}$$

Incorporating Multiple Sources of Information

- Random Variables:
 - Low Battery (B)
 - Light Off (L)
 - No Motion (N)

$$P(B|L, N) = \frac{P(L, N|B)P(B)}{P(L, N)}$$

$$P(L, N|B) = P(L|B)P(N|B)$$

Incorporating Multiple Sources of Information

- Random Variables:
 - Low Battery (B)
 - Light Off (L)
 - No Motion (N)

$$P(B|L, N) = \frac{P(L, N|B)P(B)}{P(L, N)}$$

$$P(B|L, N) = \frac{P(L|B)P(N|B)P(B)}{P(L, N)}$$

Incorporating Multiple Sources of Information

- Random Variables:
 - Low Battery (B)
 - Light Off (L)
 - No Motion (N)

$$P(L, N) = P(L|B)P(N|B)P(B) + P(L| - B)P(N| - B)P(-B)$$

Continuous Random Variables

- Example: Position of the robot in the room
- A continuous random variable possesses a probability density function (PDF), which always integrates to 1:

$$\int p(x)dx = 1$$

Normal Distribution

- Normal distributions play a major role in this lecture. We will abbreviate them as

$$N(x; \mu, \sigma^2)$$

$$p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$$

Linear Transformations

- if x and y are independent and $x \sim \mathcal{N}(\mu_x, Q_x)$, $y \sim \mathcal{N}(\mu_y, Q_y)$ with $z = Ax + By + c$, with c a constant, then:

Linear Transformations

- if x and y are independent and $x \sim \mathcal{N}(\mu_x, Q_x)$, $y \sim \mathcal{N}(\mu_y, Q_y)$ with $z = Ax + By + c$, with c a constant, then:
- z is also a normal random variable with
$$z \sim \mathcal{N}(A\mu_x + B\mu_y + c, A^T Q_x A + B^T Q_y B)$$

Linear Transformations

- if x and y are independent and $x \sim \mathcal{N}(\mu_x, Q_x)$, $y \sim \mathcal{N}(\mu_y, Q_y)$ with $z = Ax + By + c$, with c a constant, then:
- z is also a normal random variable with
$$z \sim \mathcal{N}(A\mu_x + B\mu_y + c, A^T Q_x A + B^T Q_y B)$$
- Example: Let $x \sim \mathcal{N}(0, 1)$ and $y \sim \mathcal{N}(0, 4)$ and $z = x + 2y + 3$ then:
$$z \sim \mathcal{N}(3, 1 * 1 * 1 + 2 * 4 * 2) = \mathcal{N}(3, 17)$$

Probability Axioms for Continuous Variables

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

$$p(x) = \int_y p(x|y)p(y)dy \quad (\text{continuous})$$

Probability Axioms for Continuous Variables

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

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

Probability Axioms for Continuous Variables

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

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

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

Expectation of Random Variable

$$E[X] = \sum_x xp(x) \quad (\text{discrete})$$

$$E[X] = \int_x xp(x)dx \quad (\text{continuous})$$

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

Example

- If you roll a 1 or 2, you lose \$3.
- If you roll a 3 or 4, you win \$2.
- If you roll a 5 or 6, you win \$5.

Example

- If you roll a 1 or 2, you lose \$3.
 - If you roll a 3 or 4, you win \$2.
 - If you roll a 5 or 6, you win \$5.
-
- Expected Value = (Outcome 1 or 2) * (Probability of 1 or 2) + (Outcome 3 or 4) * (Probability of 3 or 4) + (Outcome 5 or 6) * (Probability of 5 or 6)

Example

- If you roll a 1 or 2, you lose \$3.
 - If you roll a 3 or 4, you win \$2.
 - If you roll a 5 or 6, you win \$5.
-
- Expected Value = (Outcome 1 or 2) * (Probability of 1 or 2) + (Outcome 3 or 4) * (Probability of 3 or 4) + (Outcome 5 or 6) * (Probability of 5 or 6)
 - Expected Value = $(-\$3) * (1/3) + (\$2) * (1/3) + (\$5) * (1/3)$ Expected Value = $(-\$1) + (\$2/3) + (\$5/3)$
Expected Value = $\$2/3$

Example

- Let's assume that the time the battery of a robot lasts follows a probability distribution with density function: $f(x) = 0.02x$ for $0 \leq x \leq 50$, 0 otherwise. What's the expected battery life of the robot?

Example

- Let's assume that the time the battery of a robot lasts follows a probability distribution with density function: $f(x) = 0.02x$ for $0 \leq x \leq 50$, 0 otherwise. What's the expected battery life of the robot?

$$E(X) = \int_0^{50} x \cdot (0.02x) dx$$

Example

- Let's assume that the time the battery of a robot lasts follows a probability distribution with density function: $f(x) = 0.02x$ for $0 \leq x \leq 50$, 0 otherwise. What's the expected battery life of the robot?

$$E(X) = \int_0^{50} x \cdot (0.02x) dx$$

$$E(X) = 0.02 \int_0^{50} x^2 dx$$

Example

- Let's assume that the time the battery of a robot lasts follows a probability distribution with density function: $f(x) = 0.02x$ for $0 \leq x \leq 50$, 0 otherwise. What's the expected battery life of the robot?

$$E(X) = \int_0^{50} x \cdot (0.02x) dx$$

$$E(X) = 0.02 \int_0^{50} x^2 dx$$

Using the power rule of integration, we find:

$$E(X) = 0.02 \cdot \frac{x^3}{3} \Big|_0^{50}$$

Example

- Let's assume that the time the battery of a robot lasts follows a probability distribution with density function: $f(x) = 0.02x$ for $0 \leq x \leq 50$, 0 otherwise. What's the expected battery life of the robot?

$$E(X) = \int_0^{50} x \cdot (0.02x) dx$$

$$E(X) = 0.02 \int_0^{50} x^2 dx$$

Using the power rule of integration, we find:

$$E(X) = 0.02 \cdot \frac{x^3}{3} \Big|_0^{50}$$

$$E(X) = \frac{0.02}{3} \cdot (50^3 - 0^3)$$

$$E(X) = \frac{0.02}{3} \cdot 125000$$

$$E(X) = 833.33$$

Example

- Q: We Let $X \sim \text{uniform}(0,1)$. Find $E(X)$.

Example

- Q: We Let $X \sim \text{uniform}(0,1)$. Find $E(X)$.
- A: X has range $[0, 1]$ and density $f(x) = 1$. Therefore:

Example

- Q: We Let $X \sim \text{uniform}(0,1)$. Find $E(X)$.
- A: X has range $[0, 1]$ and density $f(x) = 1$. Therefore:

$$E(X) = \int_0^1 x dx = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2}$$

Example

- Q: We Let $Z \sim N(0,1)$. Find $E(Z)$.

- A: Z has range $(-\infty, \infty)$ and density $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

Therefore:

Example

- Q: We Let $Z \sim N(0,1)$. Find $E(Z)$.
- A: Z has range $(-\infty, \infty)$ and density $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

Therefore:

$$E(Z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} dz$$

Example

- Q: We Let $Z \sim N(0,1)$. Find $E(Z)$.
- A: Z has range $(-\infty, \infty)$ and density $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

Therefore:

$$E(Z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} dz = -\frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} \Bigg|_{-\infty}^{\infty} = 0$$

Random Vectors

- Normalization:
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x}) dx_1 dx_2 \cdots dx_n = 1$$

- Mean:
$$\bar{\mathbf{x}} = E[\mathbf{x}] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{x} f(\mathbf{x}) dx_1 dx_2 \cdots dx_n$$

- Covariance:

$$\Sigma = [\Sigma_{ij}] = E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T f(\mathbf{x}) dx_1 dx_2 \cdots dx_n$$

Example

- Covariance matrix of random vector: $\mathbf{x} = (x_1, x_2, x_3)$

Example

- Covariance matrix of random vector: $\mathbf{x} = (x_1, x_2, x_3)$

$$C(\mathbf{x}) = E \left\{ \begin{array}{c} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ x_3 - \mu_3 \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_2 & x_3 - \mu_3 \end{bmatrix} \end{array} \right\}$$

Example

- Covariance matrix of random vector: $\mathbf{x} = (x_1, x_2, x_3)$

$$C(\mathbf{x}) = E \left\{ \begin{array}{c} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ x_3 - \mu_3 \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_2 & x_3 - \mu_3 \end{bmatrix} \end{array} \right\}$$

$$C(\mathbf{x}) = E \left\{ \begin{array}{c} \begin{bmatrix} (x_1 - \mu_1)^2 & (x_1 - \mu_1)(x_2 - \mu_2) & (x_1 - \mu_1)(x_3 - \mu_3) \\ (x_2 - \mu_2)(x_1 - \mu_1) & (x_2 - \mu_2)^2 & (x_2 - \mu_2)(x_3 - \mu_3) \\ (x_3 - \mu_3)(x_1 - \mu_1) & (x_3 - \mu_3)(x_2 - \mu_2) & (x_3 - \mu_3)^2 \end{bmatrix} \end{array} \right\}$$

Example

- Covariance matrix of random vector: $\mathbf{x} = (x_1, x_2, x_3)$

$$C(\mathbf{x}) = E \left\{ \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ x_3 - \mu_3 \end{bmatrix} [x_1 - \mu_1 \quad x_2 - \mu_2 \quad x_3 - \mu_3] \right\}$$

$$C(\mathbf{x}) = E \left\{ \begin{bmatrix} (x_1 - \mu_1)^2 & (x_1 - \mu_1)(x_1 - \mu_2) & (x_2 - \mu_1)(x_3 - \mu_3) \\ (x_2 - \mu_2)(x_1 - \mu_1) & (x_2 - \mu_2)^2 & (x_2 - \mu_2)(x_3 - \mu_3) \\ (x_3 - \mu_3)(x_1 - \mu_1) & (x_3 - \mu_3)(x_2 - \mu_2) & (x_3 - \mu_3)^2 \end{bmatrix} \right\}$$

$$C(\mathbf{x}) = \begin{bmatrix} V(x_1) & Cov(x_1, x_2) & Cov(x_1, x_3) \\ Cov(x_1, x_2) & V(x_2) & Cov(x_2, x_3) \\ Cov(x_1, x_3) & Cov(x_2, x_3) & V(x_3) \end{bmatrix}$$

Stationarity

- Stationarity: Parameters, e.g., mean, covariance, are time-shift invariant: $E[x(t)] = \mu \forall t$
- Are the following processes z stationary:
 $z = 1 + w, w \sim \mathcal{N}(0,1)$

Stationarity

- Stationarity: Parameters, e.g., mean, covariance, are time-shift invariant: $E[x(t)] = \mu \forall t$
- Are the following processes z stationary:
 - $z = 1 + w, w \sim \mathcal{N}(0, 1)$
 - $z = 1 + w, w \sim \mathcal{N}(t, 1)$

Stationarity

- Stationarity: Parameters, e.g., mean, covariance, are time-shift invariant: $E[x(t)] = \mu \forall t$
- Are the following processes z stationary:
 - $z = 1 + w, w \sim \mathcal{N}(0,1)$
 - $z = 1 + w, w \sim \mathcal{N}(t, 1)$
 - $z = t + w, w \sim \mathcal{N}(0,1)$

Markov Chains

- A set of trials with possible outcomes, e_1, e_2, \dots is called a Markov chain if the probabilities of sampled sequences are defined by:
- $P(\{e_{j1}, \dots, e_{jT}\}) = P(e_{j1}) P(e_{j2} | e_{j1}) \dots P(e_{jT} | e_{jT-1})$

Markov Chains

- A set of trials with possible outcomes, e_1, e_2, \dots is called a Markov chain if the probabilities of sampled sequences are defined by:
- $P(\{e_{j1}, \dots, e_{jT}\}) = P(e_{j1}) P(e_{j2} | e_{j1}) \dots P(e_{jT} | e_{jT-1})$
- The probabilities of transitioning from outcome e_{k-1} to e_k are defined with a transition matrix

$$T = \begin{bmatrix} p_{11} & p_{12} & \dots \\ p_{21} & p_{22} & \dots \\ \vdots & & \\ p_{31} & p_{32} & \dots \end{bmatrix}$$

Example

- Assume three possible system states, e_1 , e_2 and e_3 :

- x_1 = Battery full
- x_2 = Battery low
- x_3 = Battery empty

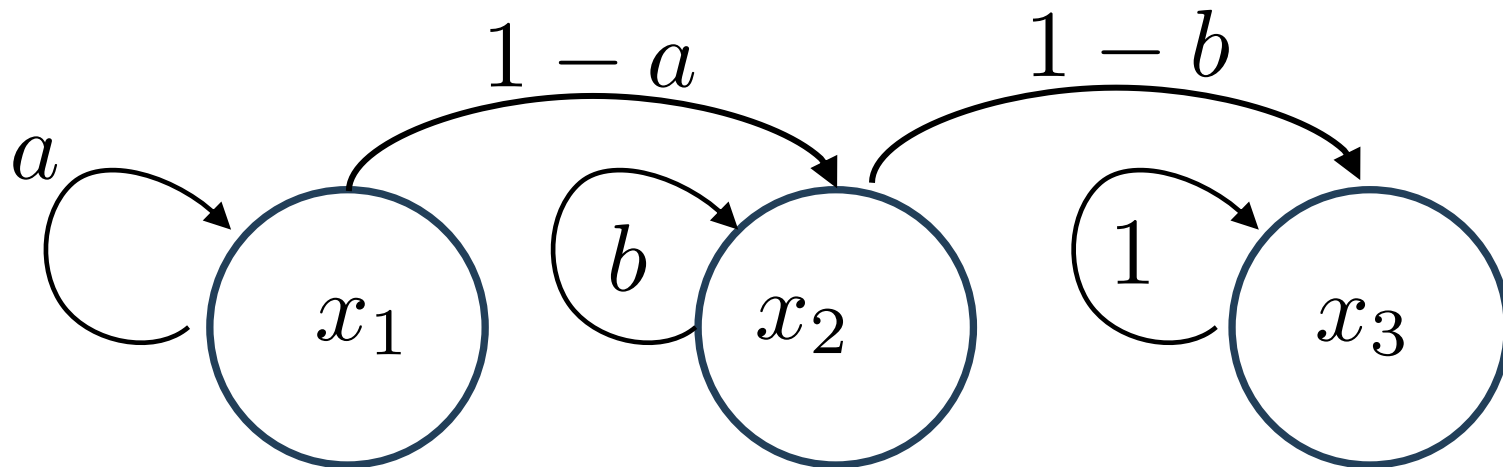
$$T = \begin{bmatrix} a & 1 - a & 0 \\ 0 & b & 1 - b \\ 0 & 0 & 1 \end{bmatrix}$$

Example

- Assume three possible system states, e_1 , e_2 and e_3 :

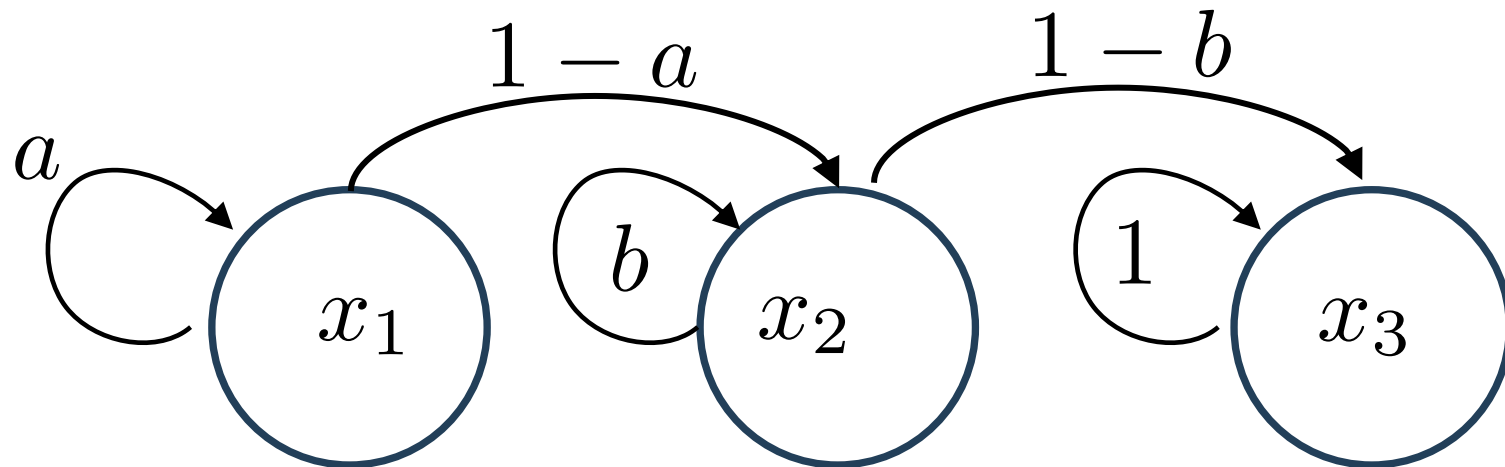
- x_1 = Battery full
- x_2 = Battery low
- x_3 = Battery empty

$$T = \begin{bmatrix} a & 1 - a & 0 \\ 0 & b & 1 - b \\ 0 & 0 & 1 \end{bmatrix}$$



Example

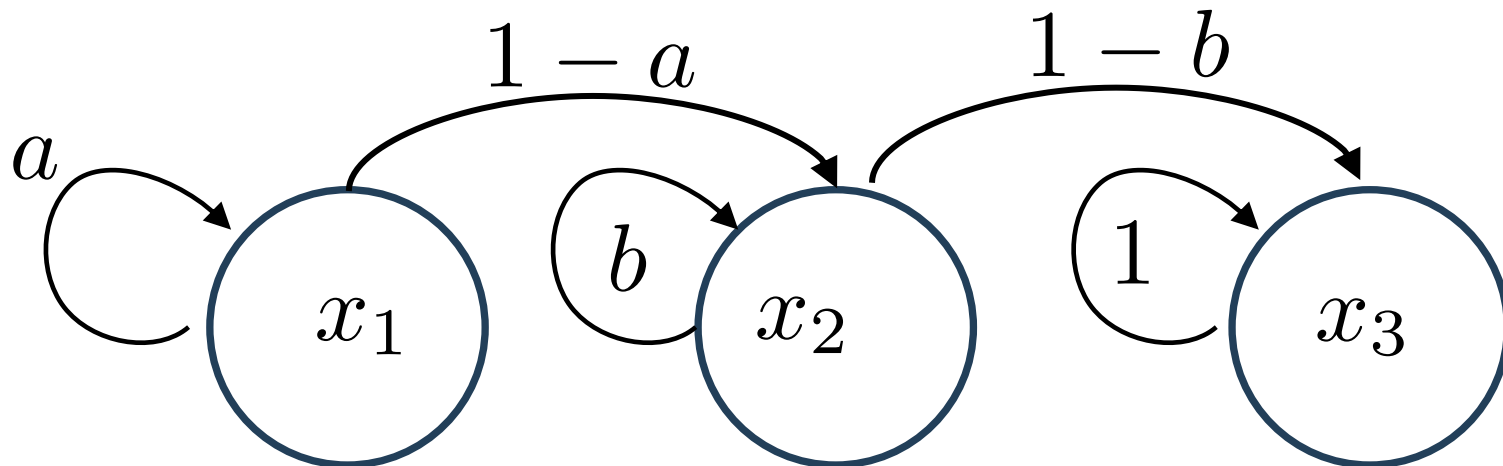
- What is the probability of the sequence (“battery full”, “battery low”, “battery low”) given that the robot starts with battery full?



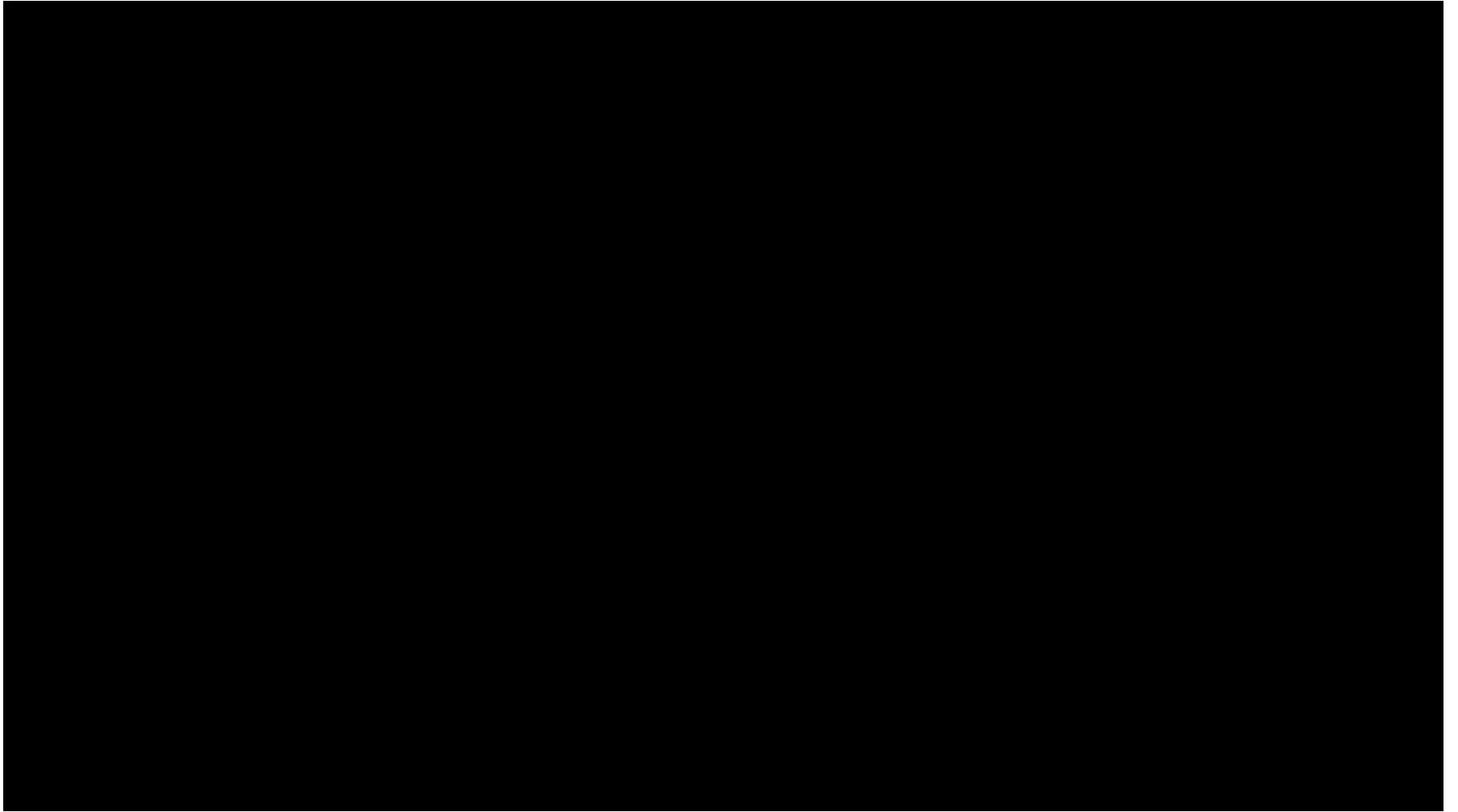
Example

- What is the probability of the sequence {"battery full", "battery low", "battery low"} given that the robot starts with battery full?

$$\begin{aligned} P(\{x_1, x_2, x_2\}) &= P(x_1)P(x_2|x_1)P(x_2|x_2) \\ &= 1 * (1 - a) * b \end{aligned}$$



Reasoning over uncertainty



State

- A *state* captures all the information that the robot needs to make decisions. It may change over time such as its location.
- There are two types of interaction that the robot can have with the environment.

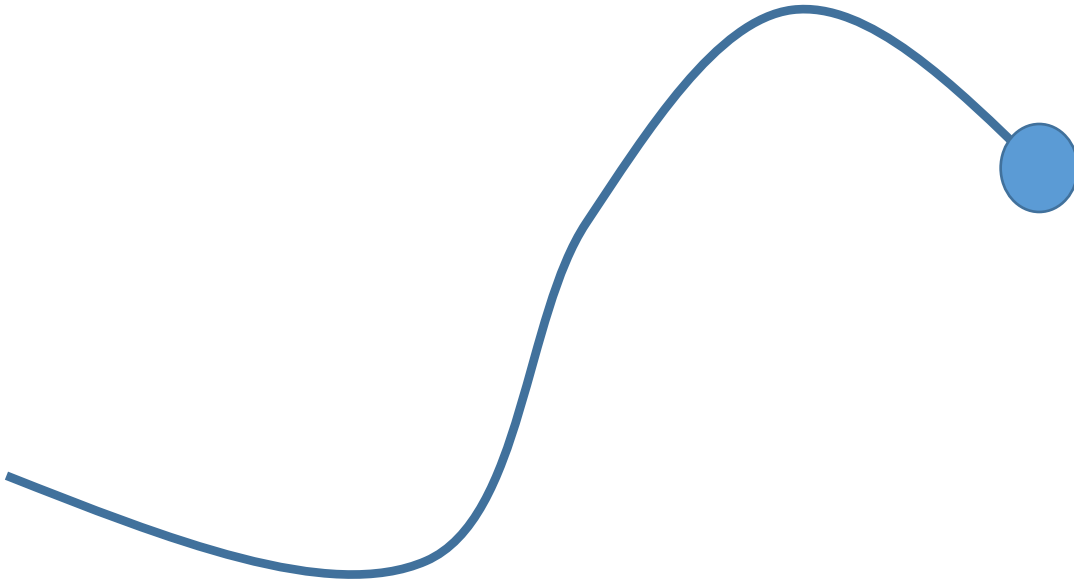
Robot Interaction

- Robot is at a state x_t
- Takes a measurement z_t
- Takes action u_t
- Transitions to a new state x_{t+1}

Actuation

$$p(x_t | x_{1:t-1}, z_{1:t-1}, u_{1:t}) = p(x_t | x_{t-1}, u_t)$$

(Markov assumption)



Sensing

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

(Markov assumption)