CS 287 Lecture 11 (Fall 2019)

Probability Review, Bayes Filters, Gaussians

Pieter Abbeel

UC Berkeley EECS

Many slides adapted from Thrun, Burgard and Fox, Probabilistic Robotics

Outline

- Probability Review
- Bayes Filters
- Gaussians

Why probability in robotics?

- Often the state of the robot and of its environment are unknown and only noisy sensors are available
  - Probability provides a framework to fuse sensory information
  - → Result: probability distribution over possible states of robot and environment
- Dynamics is often stochastic, hence can't optimize for a particular outcome, but only optimize to obtain a good distribution over outcomes
  - Probability provides a framework to reason in this setting
  - → Ability to find good control policies for stochastic dynamics and environments

Example 1: Helicopter

- State: position, orientation, velocity, angular rate
- Sensors:
  - GPS : noisy estimate of position (sometimes also velocity)
  - Inertial sensing unit: noisy measurements from
    (i) 3-axis gyro [=angular rate sensor],
    (ii) 3-axis accelerometer [measures acceleration + gravity; e.g., measures (0,0,0) in free-fall],
    (iii) 3-axis magnetometer
- Dynamics:
  - Noise from: wind, unmodeled dynamics in engine, servos, blades

Example 2: Mobile robot inside building

- State: position and heading
- Sensors:
  - Odometry (=sensing motion of actuators): e.g., wheel encoders
  - Laser range finder:
    - Measures time of flight of a laser beam between departure and return
    - Return is typically happening when hitting a surface that reflects the beam back to where it came from
- Dynamics:
  - Noise from: wheel slippage, unmodeled variation in floor

Outline

- ***Probability Review***
- Bayes Filters
- Gaussians

Axioms of Probability Theory

$$0 \le \Pr(A) \le 1$$

$$\Pr(\Omega) = 1 \qquad \Pr(\phi) = 0$$

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

Pr*(A)* denotes probability that the outcome ω is an element of the set of possible outcomes *A*. *A* is often called an event. Same for *B.*

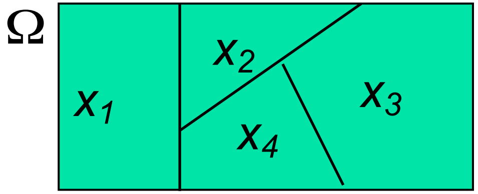
Ω is the set of all possible outcomes.

φ is the empty set.

Using the Axioms

$$\begin{aligned}
\Pr(A \cup (\Omega \setminus A)) &= \Pr(A) + \Pr(\Omega \setminus A) - \Pr(A \cap (\Omega \setminus A)) \\
\Pr(\Omega) &= \Pr(A) + \Pr(\Omega \setminus A) - \Pr(\phi) \\
1 &= \Pr(A) + \Pr(\Omega \setminus A) - 0 \\
\Pr(\Omega \setminus A) &= 1 - \Pr(A)
\end{aligned}$$

Discrete Random Variables



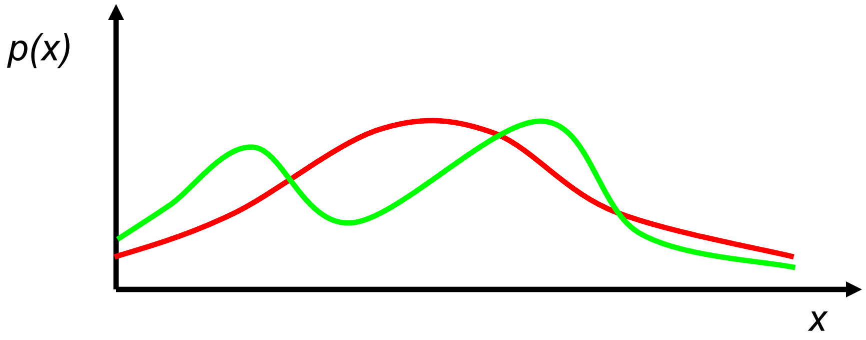
- *X* denotes a random variable.
- *X* can take on a countable number of values in $\{x_1, x_2, \ldots, x_n\}$.
- $P(X=x_i)$, or $P(x_i)$, is the probability that the random variable *X* takes on value x_i.
- $P(.)$ is called probability mass function.
- *E.g., X models the outcome of a coin flip,* x_1 = head, x_2 = tail, $P(x_1) = 0.5$, $P(x_2) = 0.5$

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.



Joint and Conditional Probability

- $P(X=x \text{ and } Y=y) = P(x,y)$

- X and Y are independent iff

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

- $P(x \mid y)$ is the probability of x given y

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

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

- If X and Y are independent then

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

- *Same for probability densities, just $P \rightarrow p$*

Law of Total Probability, Marginals

Discrete case

$$\sum_x P(x) = 1$$

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

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

Continuous case

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

$$p(x) = \int p(x, y)\, dy$$

$$p(x) = \int p(x \mid y) p(y)\, dy$$

Bayes Rule

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

$$\Rightarrow$$

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

Normalization

$$P(x|y) = \frac{P(y|x)\ P(x)}{P(y)} = \eta\ P(y|x)\, P(x)$$

$$\eta = P(y)^{-1} = \frac{1}{\sum_x P(y|x) P(x)}$$

Algorithm:

$$\forall x : \mathrm{aux}_{x|y} = P(y|x)\ P(x)$$

$$\eta = \frac{1}{\sum_x \mathrm{aux}_{x|y}}$$

$$\forall x : P(x|y) = \eta\ \mathrm{aux}_{x|y}$$

Conditioning

- Law of total probability:

$$P(x) = \int P(x, z)dz$$

$$P(x) = \int P(x \mid z)P(z)dz$$

$$P(x \mid y) = \int P(x \mid y, z)\, P(z \mid y)\, dz$$

Bayes Rule with Background Knowledge

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

Conditional Independence

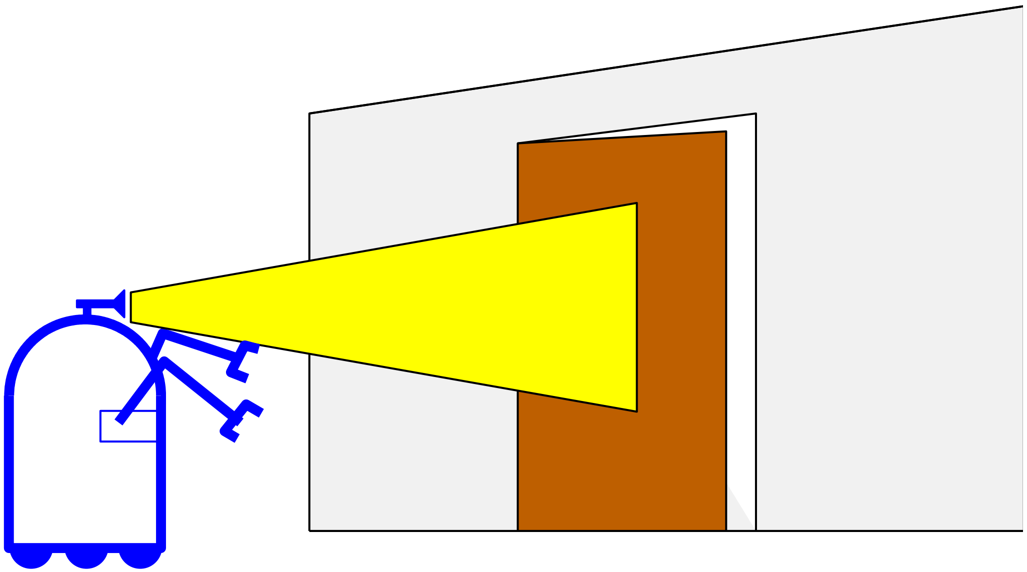
$$P(x, y \mid z) = P(x \mid z) P(y \mid z)$$

equivalent to $P(x \mid z) = P(x \mid z, y)$

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

Simple Example of State Estimation

- Suppose a robot obtains measurement z
- What is $P(open|z)$?



Causal vs. Diagnostic Reasoning

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

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

Example

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

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

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

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

- z raises the probability that the door is open.

Combining Evidence

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

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*.

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

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

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

Example: Second Measurement

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

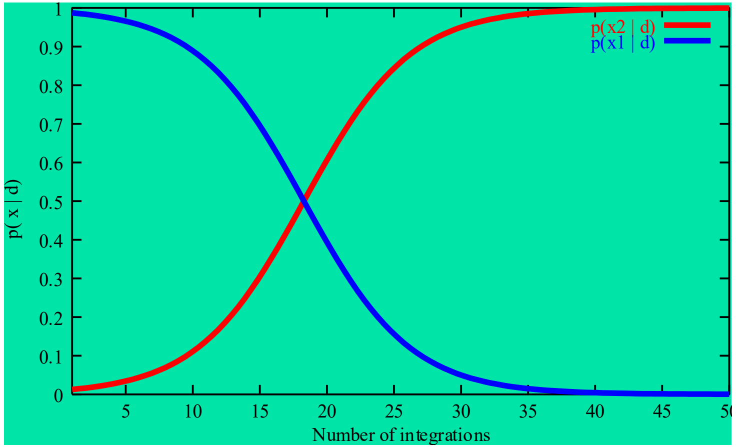
$$P(open \mid z_2, z_1) = \frac{P(z_2 \mid open)\, P(open \mid z_1)}{P(z_2 \mid open)\, P(open \mid z_1) + P(z_2 \mid \neg open)\, P(\neg open \mid z_1)}$$

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

- z_2 lowers the probability that the door is open.

A Typical Pitfall

- Two possible locations x_1 and x_2
- $P(x_1)=0.99$
- $P(z|x_2)=0.09$
  $P(z|x_1)=0.07$



If measurements are not independent but are treated as independent
→ can quickly end up overconfident

Outline

- Probability Review
- ***Bayes Filters***
- Gaussians

Actions

- Often the world is **dynamic** since
  - **actions carried out by the robot**,
  - **actions carried out by other agents**,
  - or just the **time** passing by

  change the world.

- How can we **incorporate** such **actions**?

Typical Actions

- The robot **turns its wheels** to move
- The robot **uses its manipulator** to grasp an object
- Plants grow over **time**…

- Actions are **never carried out with absolute certainty**.
- In contrast to measurements, **actions generally increase the uncertainty**.

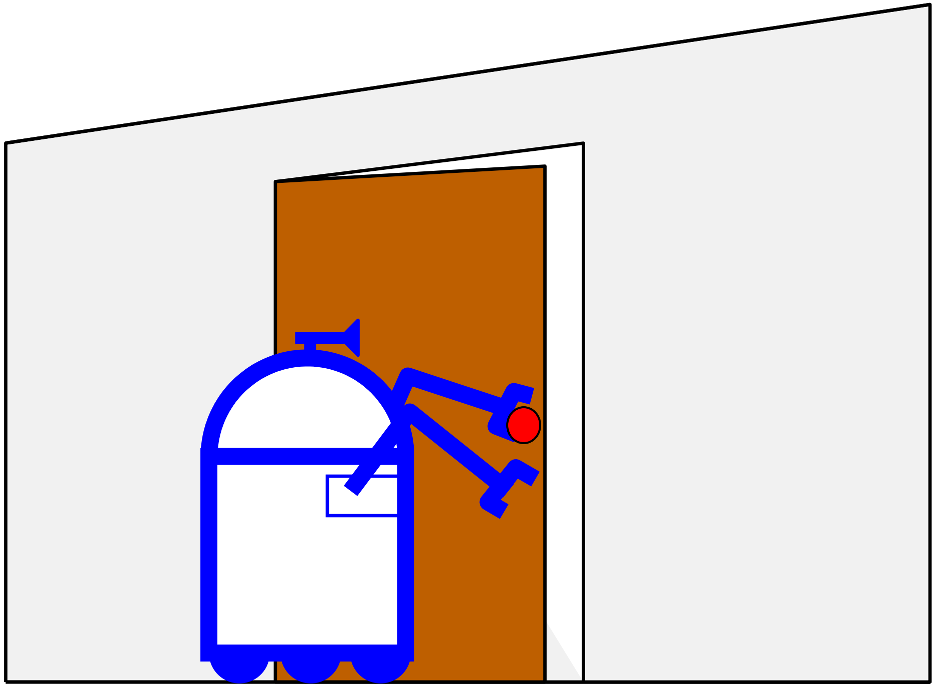
Modeling Actions

- To incorporate the outcome of an action *u* into the current “belief”, we use the conditional pdf

$$P(x'|u,x)$$

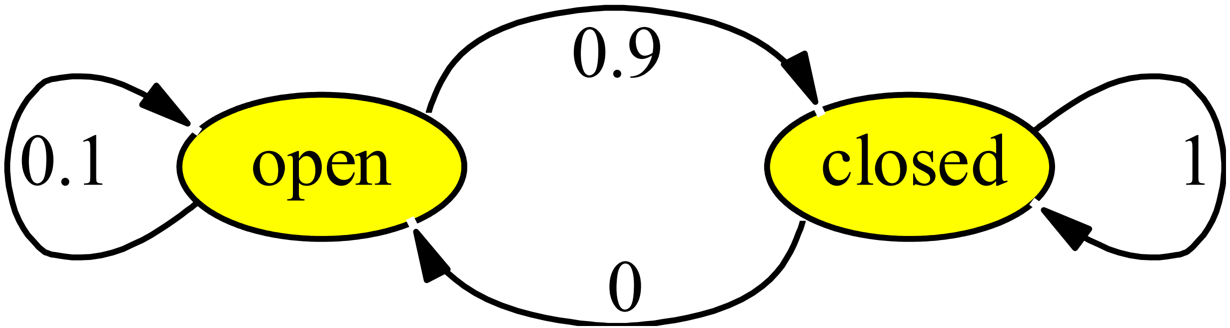
- This term specifies the pdf that **executing *u* changes the state from *x* to *x'***.

Example: Closing the door



State Transitions

$P(x'|u,x)$ for u = “close door”:



If the door is open, the action “close door” succeeds in 90% of all cases.

Integrating the Outcome of Actions

Continuous case:

$$P(x' \mid u) = \int P(x' \mid u, x) P(x)\, dx$$

Discrete case:

$$P(x' \mid u) = \sum P(x' \mid u, x) P(x)$$

Example: The Resulting Belief

$$P(closed \mid u) = \sum P(closed \mid u, x) P(x)$$

$$= P(closed \mid u, open) P(open)$$

$$+ P(closed \mid u, closed) P(closed)$$

$$= \frac{9}{10} * \frac{5}{8} + \frac{1}{1} * \frac{3}{8} = \frac{15}{16}$$

$$P(open \mid u) = \sum P(open \mid u, x) P(x)$$

$$= P(open \mid u, open) P(open)$$

$$+ P(open \mid u, closed) P(closed)$$

$$= \frac{1}{10} * \frac{5}{8} + \frac{0}{1} * \frac{3}{8} = \frac{1}{16}$$

$$= 1 - P(closed \mid u)$$

Measurements

- Bayes rule

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

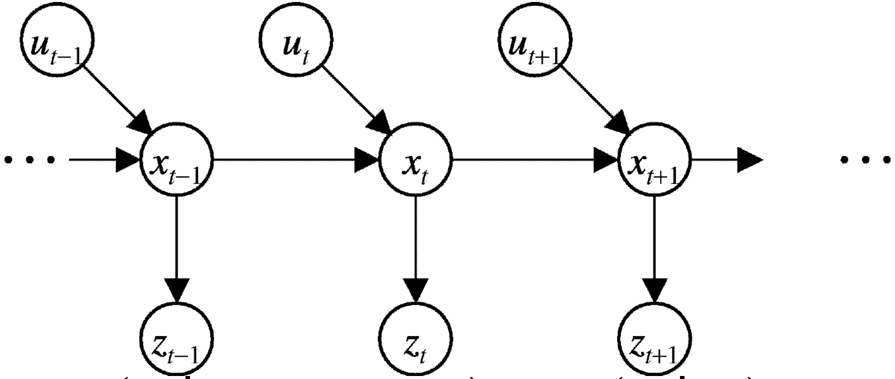
Bayes Filters: Framework

- **Given:**
  - Stream of observations *z* and action data *u:*

    $$d_t = \{u_1, z_1 \ldots, u_t, z_t\}$$

  - Sensor model *P(z|x).*
  - Action model *P(x'|u,x)*.
  - Prior probability of the system state *P(x).*
- **Wanted:**
  - Estimate of the state *X* of a dynamical system.
  - The posterior of the state is also called **Belief**: $Bel(x_t) = P(x_t \mid u_1, z_1 \ldots, u_t, z_t)$

Markov Assumption



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

Underlying Assumptions

- Static world
- Independent noise
- Perfect model, no approximation errors

Bayes Filters

z = observation
u = action
x = state

$$Bel(x_t) = P(x_t \mid u_1, z_1 \ldots, u_t, z_t)$$

Bayes $$= \eta \; P(z_t \mid x_t, u_1, z_1, \ldots, u_t) \; P(x_t \mid u_1, z_1, \ldots, u_t)$$

Markov $$= \eta \; P(z_t \mid x_t) \; P(x_t \mid u_1, z_1, \ldots, u_t)$$

Total prob. $$= \eta \; P(z_t \mid x_t) \int P(x_t \mid u_1, z_1, \ldots, u_t, x_{t-1}) \; P(x_{t-1} \mid u_1, z_1, \ldots, u_t) \; dx_{t-1}$$

Markov $$= \eta \; P(z_t \mid x_t) \int P(x_t \mid u_t, x_{t-1}) \; P(x_{t-1} \mid u_1, z_1, \ldots, u_t) \; dx_{t-1}$$

Markov $$= \eta P(z_t \mid x_t) \int P(x_t \mid u_t, x_{t-1}) P(x_{t-1} \mid u_1, z_1, \ldots, z_{t-1}) \; dx_{t-1}$$

$$= \eta \; P(z_t \mid x_t) \int P(x_t \mid u_t, x_{t-1}) \; Bel(x_{t-1}) \; dx_{t-1}$$

Bayes Filters

$$Bel(x_t) = \eta\ P(z_t \mid x_t) \int P(x_t \mid u_t, x_{t-1})\ Bel(x_{t-1})\ dx_{t-1}$$

$\eta = 0$

If *d* is a perceptual data item *z* then

For all *x* do

$$Bel'(x) = P(z \mid x) Bel(x)$$

$$\eta = \eta + Bel'(x)$$

For all *x* do

$$Bel'(x) = \eta^{-1} Bel'(x)$$

Else if *d* is an action data item *u* then

For all *x* do

$$Bel'(x) = \int P(x \mid u, x')\ Bel(x')\ dx'$$

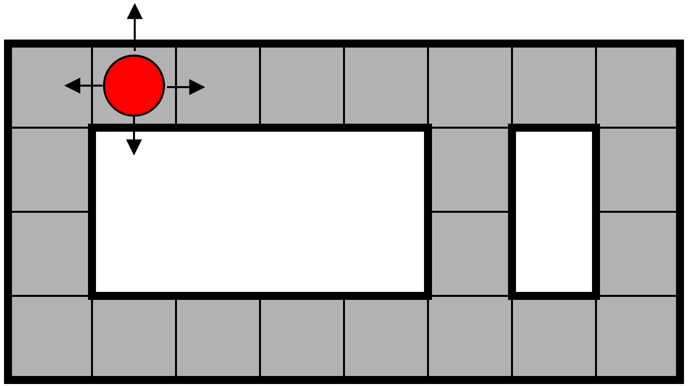
Return *Bel'(x)*

Summary

- Bayes rule allows us to compute probabilities that are hard to assess otherwise.
- Under the Markov assumption, recursive Bayesian updating can be used to efficiently combine evidence.
- Bayes filters are a probabilistic tool for estimating the state of dynamic systems.

Example: Robot Localization

Example from Michael Pfeiffer



Prob

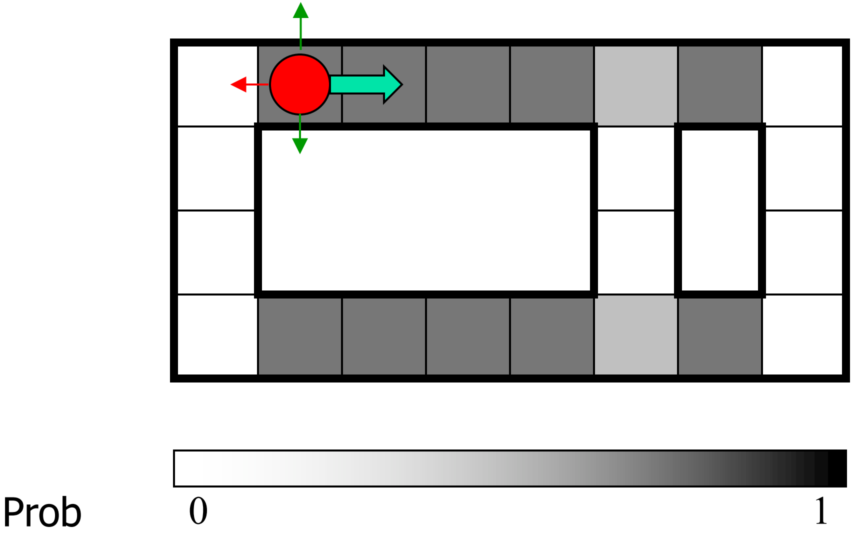
t=0

Sensor model: never more than 1 mistake

Know the heading (North, East, South or West)

Motion model: may not execute action with small prob.

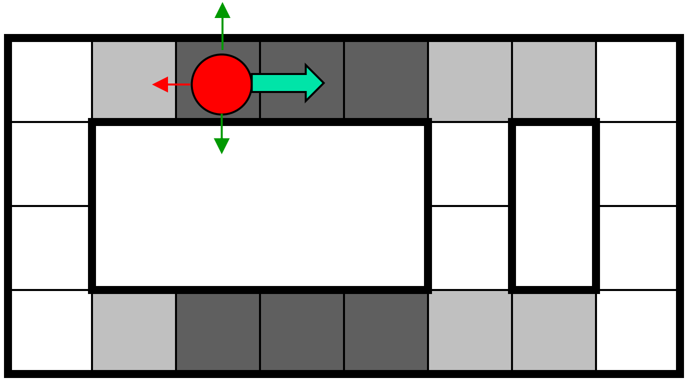
Example: Robot Localization

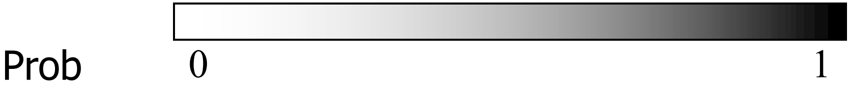


t=1

Lighter grey: was possible to get the reading, but less likely b/c required 1 mistake

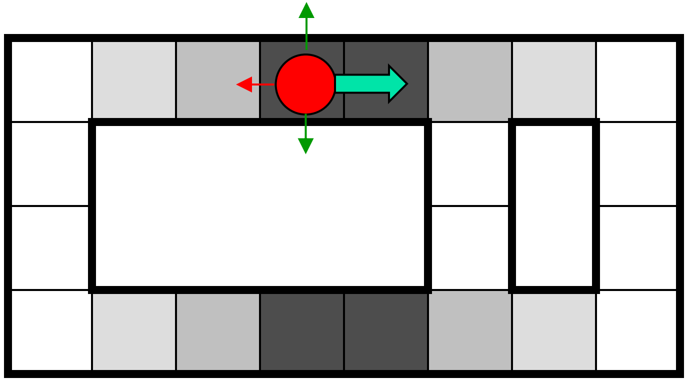
Example: Robot Localization

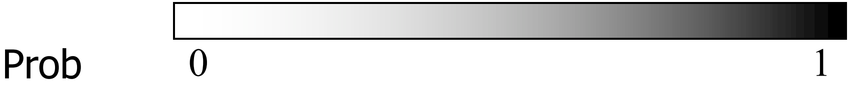




t=2

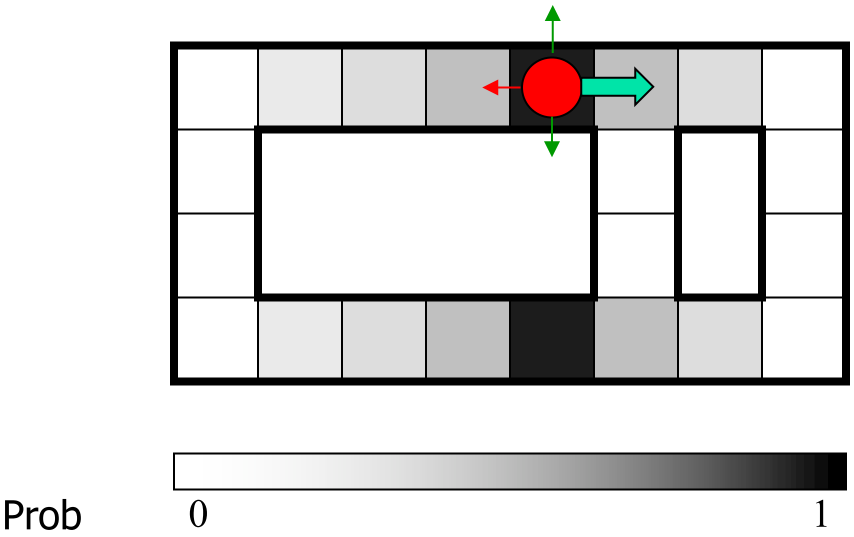
Example: Robot Localization





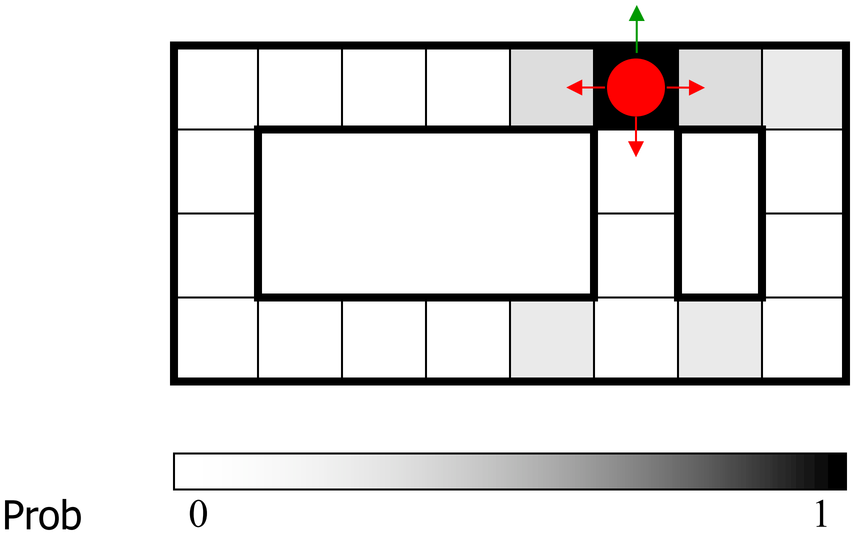
t=3

Example: Robot Localization



t=4

Example: Robot Localization



t=5

Outline

- Probability Review
- Bayes Filters
- ***Gaussians***

Gaussians -- Outline

- Univariate Gaussian
- Multivariate Gaussian
- Law of Total Probability
- Conditioning (Bayes' rule)

Disclaimer: lots of linear algebra in next few lectures. See course homepage for pointers for brushing up your linear algebra.

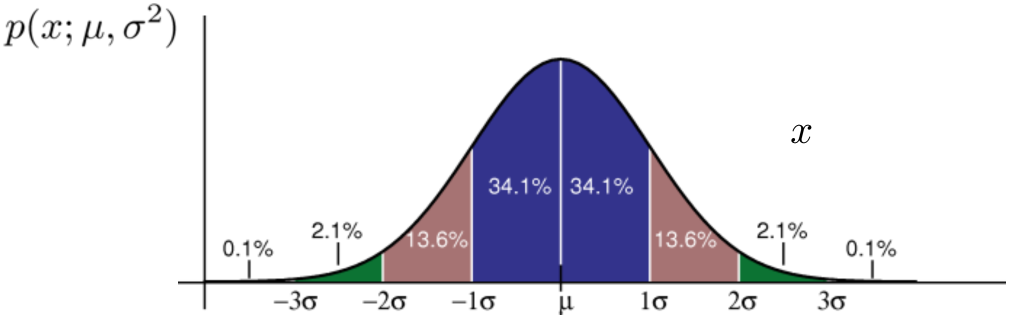
In fact, pretty much all computations with Gaussians will be reduced to linear algebra!

Univariate Gaussian

- Gaussian distribution with mean μ, and standard deviation σ:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

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



$X \sim \mathcal{N}(\mu, \sigma^2)$

Properties of Gaussians

$p(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$

- Densities integrate to one: $\int_{-\infty}^{\infty} p(x; \mu, \sigma^2) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx = 1$

- Mean:

$$\begin{aligned} \mathrm{E}_X[X] &= \int_{-\infty}^{\infty} x p(x; \mu, \sigma^2) dx \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx \\ &= \mu \end{aligned}$$

- Variance:

$$\begin{aligned} \mathrm{E}_X[(X-\mu)^2] &= \int_{-\infty}^{\infty} (x-\mu)^2 p(x; \mu, \sigma^2) dx \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx \\ &= \sigma^2 \end{aligned}$$

Central limit theorem (CLT)

- Classical CLT:
  - Let X_1, X_2, … be an infinite sequence of *independent* random variables with $E\, X_i = \mu$, $E(X_i - \mu)^2 = \sigma^2$
  - Define $Z_n = ((X_1 + \ldots + X_n) - n\,\mu) / (\sigma\, n^{1/2})$
  - Then for the limit of n going to infinity we have that Z_n is distributed according to N(0,1)
- Crude statement: random variables that result from the addition of lots of small effects are well captured by a Gaussian.

Multivariate Gaussians

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$$

$$\int \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right) dx = 1$$

For a matrix $A \in \mathbb{R}^{n\times n}$, $|A|$ denotes the determinant of A.

For a matrix $A \in \mathbb{R}^{n\times n}$, A^{-1} denotes the inverse of A, which satisfies $A^{-1}A = I = AA^{-1}$ with $I \in \mathbb{R}^{n\times n}$ the identity matrix with all diagonal entries equal to one, and all off-diagonal entries equal to zero.

Hint: often when trying to understand matrix equations, it's easier to first consider the special case of the dimensions of the matrices being one-by-one. Once parsing them that way makes sense, a good second step can be to parse them assuming all matrices are diagonal matrices. Once parsing them that way makes sense, usually it is only a small step to understand the general case.

Multivariate Gaussians

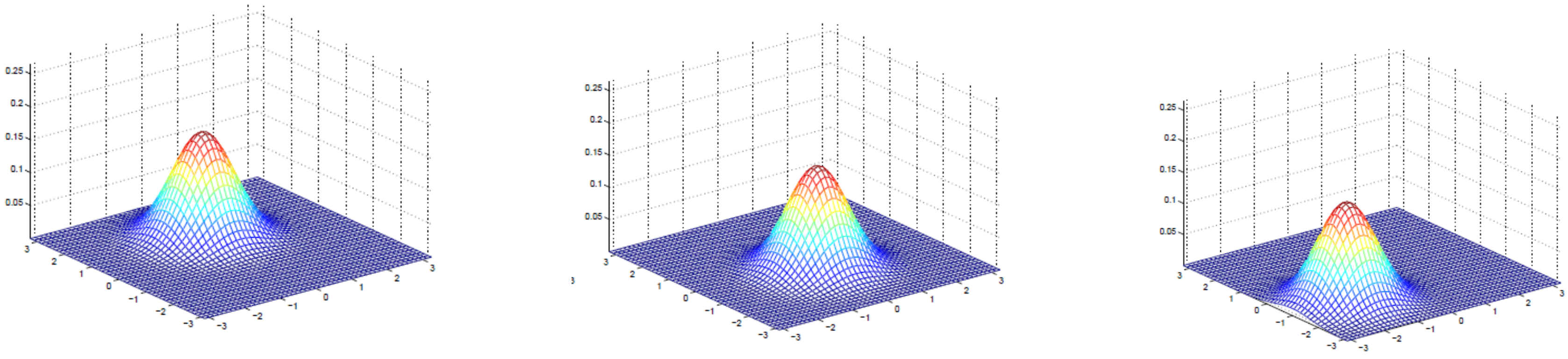
$$\mathsf{E}_X[X_i] = \int x_i p(x;\mu,\Sigma)dx = \mu_i$$
$$\mathsf{E}_X[X] = \int x p(x;\mu,\Sigma)dx = \mu$$

(integral of vector = vector of integrals of each entry)

$$\mathsf{E}_X[(X_i-\mu_i)(X_j-\mu_j)] = \int (x_i-\mu_i)(x_j-\mu_j)p(x;\mu,\Sigma)dx = \Sigma_{ij}$$
$$\mathsf{E}_X[(X-\mu)(X-\mu)^\top] = \int [(X-\mu)(X-\mu)^\top p(x;\mu,\Sigma)dx = \Sigma$$

(integral of matrix = matrix of integrals of each entry)

Multivariate Gaussians: Examples

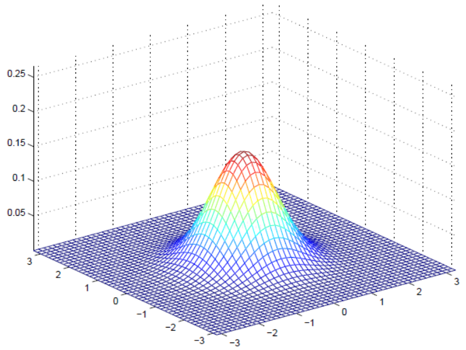


- μ = [1; 0]
- Σ = [1 0; 0 1]

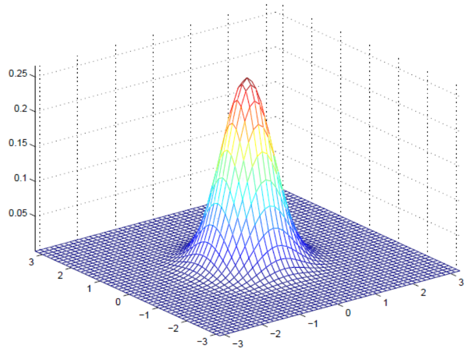
- μ = [-.5; 0]
- Σ = [1 0; 0 1]

- μ = [-1; -1.5]
- Σ = [1 0; 0 1]

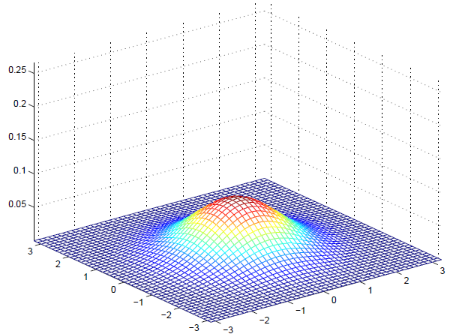
Multivariate Gaussians: Examples



- μ = [0; 0]
- Σ = [1 0 ; 0 1]

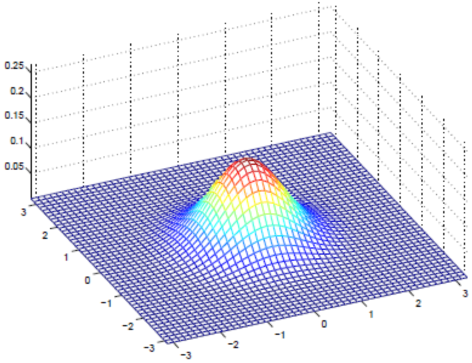


- μ = [0; 0]
- Σ = [.6 0 ; 0 .6]

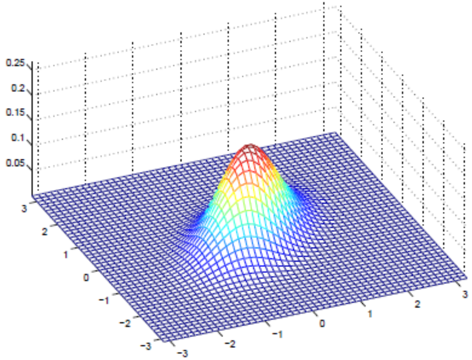


- μ = [0; 0]
- Σ = [2 0 ; 0 2]

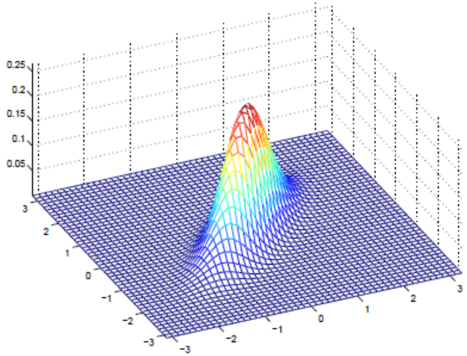
Multivariate Gaussians: Examples



- $\mu = [0; 0]$
- $\Sigma = [1\ \ 0; 0\ \ 1]$

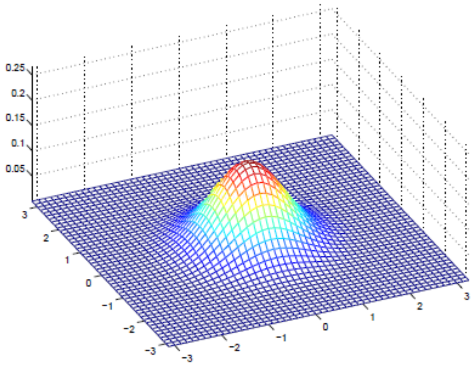


- $\mu = [0; 0]$
- $\Sigma = [1\ \ 0.5; 0.5\ 1]$

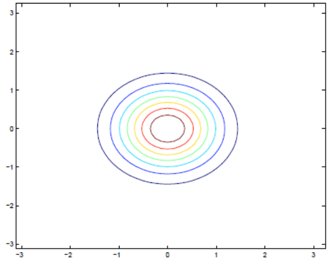


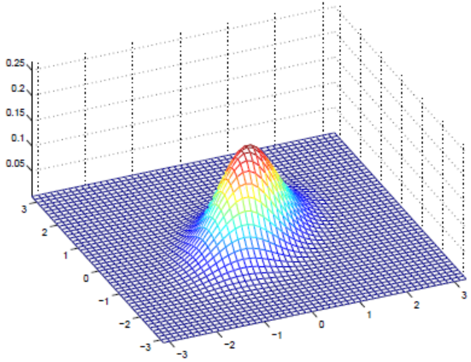
- $\mu = [0; 0]$
- $\Sigma = [1\ \ 0.8; 0.8\ \ 1]$

Multivariate Gaussians: Examples

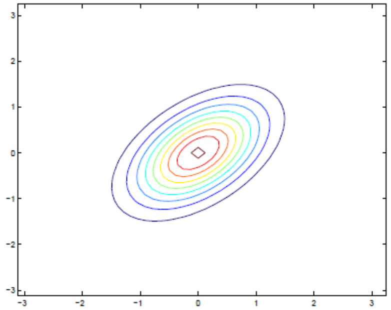


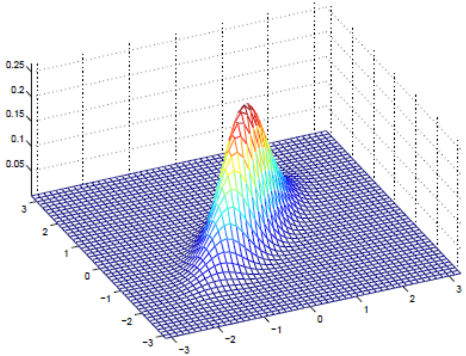
- $\mu = [0; 0]$
- $\Sigma = [1 \;\; 0; 0 \;\; 1]$



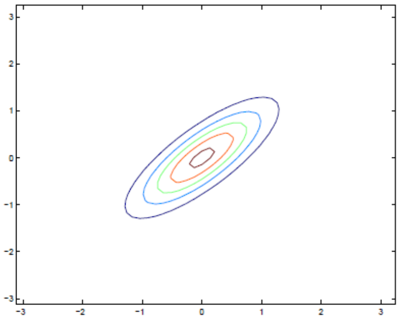


- $\mu = [0; 0]$
- $\Sigma = [1 \;\; 0.5; 0.5 \;\; 1]$

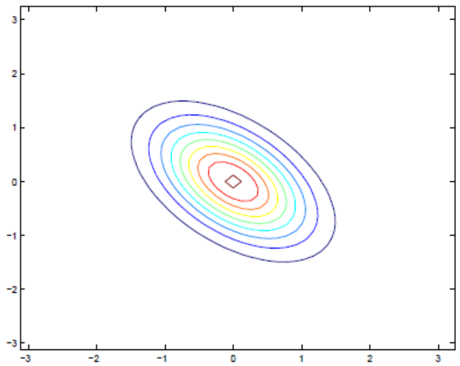




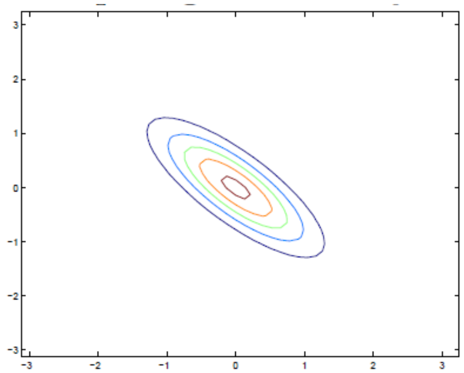
- $\mu = [0; 0]$
- $\Sigma = [1 \;\; 0.8; 0.8 \;\; 1]$



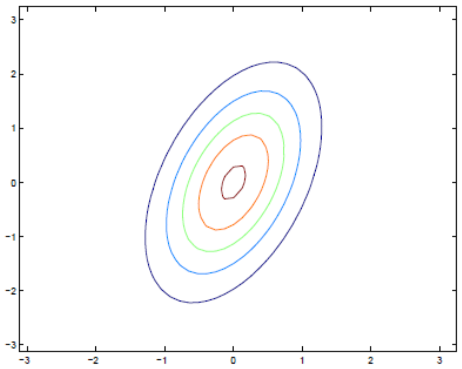
Multivariate Gaussians: Examples



- μ = [0; 0]
- Σ = [1 -0.5 ; -0.5 1]



- μ = [0; 0]
- Σ = [1 -0.8 ; -0.8 1]



- μ = [0; 0]
- Σ = [3 0.8 ; 0.8 1]

Partitioned Multivariate Gaussian

- Consider a multi-variate Gaussian and partition random vector into (X, Y).

$$\mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$$

$$p(\begin{bmatrix}x \\ y\end{bmatrix}; \mu, \Sigma) = \frac{1}{(2\pi)^{(n/2)}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix}x \\ y\end{bmatrix} - \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}\right)^\top \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}^{-1} \left(\begin{bmatrix}x \\ y\end{bmatrix} - \begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}\right)\right)$$

$$\begin{aligned}
\mu_X &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[X] \\
\mu_Y &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[Y] \\
\Sigma_{XX} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(X-\mu_X)(X-\mu_X)^\top] \\
\Sigma_{YY} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(Y-\mu_Y)(Y-\mu_Y)^\top] \\
\Sigma_{XY} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(X-\mu_X)(Y-\mu_Y)^\top] = \Sigma_{YX}^\top \\
\Sigma_{YX} &= \mathrm{E}_{(X,Y)\sim\mathcal{N}(\mu,\Sigma)}[(Y-\mu_Y)(X-\mu_X)^\top] = \Sigma_{XY}^\top
\end{aligned}$$

Partitioned Multivariate Gaussian: Dual Representation

- Precision matrix $$\Gamma = \Sigma^{-1} = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}^{-1} = \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix} \quad (1)$$

$$p(\begin{bmatrix} x \\ y \end{bmatrix}; \mu, \Sigma) = \frac{1}{(2\pi)^{(n/2)}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}\right)^\top \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}\right)\right)$$

- Straightforward to verify from (1) that:

$$\begin{aligned}
\Sigma_{XX} &= \left(\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}\right)^{-1} \\
\Sigma_{YY} &= \left(\Gamma_{YY} - \Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XY}\right)^{-1} \\
\Sigma_{XY} &= -\Gamma_{XX}^{-1}\Gamma_{XY}\left(\Gamma_{YY} - \Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XY}\right)^{-1} = \Sigma_{YX}^\top \\
\Sigma_{YX} &= -\Gamma_{YY}^{-1}\Gamma_{YX}\left(\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}\right)^{-1} = \Sigma_{XY}^\top
\end{aligned}$$

- And swapping the roles of Sigma and Gamma:

$$\begin{aligned}
\Gamma_{XX} &= \left(\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right)^{-1} \\
\Gamma_{YY} &= \left(\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)^{-1} \\
\Gamma_{XY} &= -\Sigma_{XX}^{-1}\Sigma_{XY}\left(\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)^{-1} = \Gamma_{YX}^\top \\
\Gamma_{YX} &= -\Sigma_{YY}^{-1}\Sigma_{YX}\left(\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right)^{-1} = \Gamma_{XY}^\top
\end{aligned}$$

Marginalization: p(x) = ?

$$p(\begin{bmatrix} x \\ y \end{bmatrix}; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}\right)^\top \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}\right)\right)$$

We integrate out over y to find the marginal:

$$\begin{aligned}
p(x) &= \int p(\begin{bmatrix} x \\ y \end{bmatrix}; \mu, \Sigma) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) + (y-\mu_Y)^\top \Gamma_{YY}(y-\mu_Y) + 2(y-\mu_Y)^\top \Gamma_{YX}(x-\mu_X)\right)\right) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) + (y-\mu_Y)^\top \Gamma_{YY}(y-\mu_Y) + 2(y-\mu_Y)^\top \Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X) + (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) \int \exp\left(-\frac{1}{2}\left((y-\mu_Y)^\top \Gamma_{YY}(y-\mu_Y) + 2(y-\mu_Y)^\top \Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X) + (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) \int \exp\left(-\frac{1}{2}\left((y-\mu_Y+\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X))^\top \Gamma_{YY}(y-\mu_Y+\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X))\right)\right) dy \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) (2\pi)^{n_Y/2}|\Gamma_{YY}^{-1}|^{1/2} \\
&= \frac{(2\pi)^{n_Y/2}|\Gamma_{YY}^{-1}|^{1/2}}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}(x-\mu_X)\right)\right) \\
&= \frac{(2\pi)^{n_Y/2}|\Gamma_{YY}^{-1}|^{1/2}}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left((x-\mu_X)^\top \left(\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX}\right)(x-\mu_X)\right)\right)
\end{aligned}$$

Hence we have:

$$X \sim \mathcal{N}(\mu_X, (\Gamma_{XX} - \Gamma_{XY}\Gamma_{YY}^{-1}\Gamma_{YX})^{-1}) = \mathcal{N}(\mu_X, \Sigma_{XX})$$

Note: ***if** we had known beforehand* that p(x) would be a Gaussian distribution, then we could have found the result more quickly. We would have just needed to find $\mu_X = \mathrm{E}[X]$ and $\Sigma_{XX} = \mathrm{E}[(X-\mu_X)(X-\mu_X)^\top]$, which we had available through $\mathcal{N}(\mu, \Sigma)$

Marginalization Recap

If

$$(X, Y) \sim \mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$$

Then

$$\begin{aligned} X &\sim \mathcal{N}(\mu_X, \Sigma_{XX}) \\ Y &\sim \mathcal{N}(\mu_Y, \Sigma_{YY}) \end{aligned}$$

Self-quiz

Test your understanding of the completion of squares trick! Let $A \in \mathbf{R}^{n\times n}$ be a positive definite matrix, $b \in \mathbf{R}^n$, and $c \in \mathbf{R}$. Prove that

$$\int_{x\in\mathbf{R}^n} \exp\left(-\frac{1}{2}x^T Ax - x^T b - c\right) dx$$

$$= \frac{(2\pi)^{n/2}}{|A|^{1/2}\exp(c - \frac{1}{2}b^T A^{-1} b)}.$$

Conditioning: $p(x \mid Y = y_0) = ?$

$$p\left(\begin{bmatrix} x \\ y \end{bmatrix}; \mu, \Sigma\right) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}\right)^\top \begin{bmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}\right)\right)$$

We have

$$\begin{aligned}
p(x|Y=y_0) &\propto p\left(\begin{bmatrix} x \\ y_0 \end{bmatrix}; \mu, \Sigma\right) \\
&\propto \exp\left(-\frac{1}{2}(x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}(y_0-\mu_Y) - \frac{1}{2}(y_0-\mu_Y)^\top \Gamma_{YY}(y_0-\mu_Y)\right) \\
&\propto \exp\left(-\frac{1}{2}(x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XY}(y_0-\mu_Y)\right) \\
&= \exp\left(-\frac{1}{2}(x-\mu_X)^\top \Gamma_{XX}(x-\mu_X) - (x-\mu_X)^\top \Gamma_{XX}\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y) - \frac{1}{2}(y_0-\mu_Y)\Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XX}\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y) + \frac{1}{2}(y_0-\mu_Y)\Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XX}\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y)\right) \\
&= \exp\left(-\frac{1}{2}(x-\mu_X+\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y))^\top \Gamma_{XX}(x-\mu_X+\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y))\right) \exp\left(\frac{1}{2}(y_0-\mu_Y)\Gamma_{YX}\Gamma_{XX}^{-1}\Gamma_{XX}\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y)\right) \\
&\propto \exp\left(-\frac{1}{2}(x-\mu_X+\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y))^\top \Gamma_{XX}(x-\mu_X+\Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y))\right)
\end{aligned}$$

Hence we have:

$$\begin{aligned}
X|Y=y_0 &\sim \mathcal{N}(\mu_X - \Gamma_{XX}^{-1}\Gamma_{XY}(y_0-\mu_Y), \Gamma_{XX}^{-1}) \\
&= \mathcal{N}(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y_0-\mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX})
\end{aligned}$$

- Conditional mean moved according to correlation and variance on measurement
- Conditional covariance does not depend on y_0

Conditioning Recap

If

$$(X, Y) \sim \mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{bmatrix}\right)$$

Then

$$\begin{aligned} X|Y = y_0 &\sim \mathcal{N}(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y_0 - \mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}) \\ Y|X = x_0 &\sim \mathcal{N}(\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(x_0 - \mu_X), \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}) \end{aligned}$$