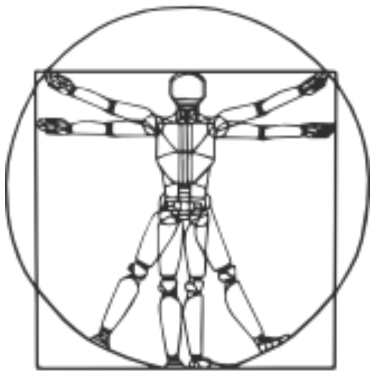
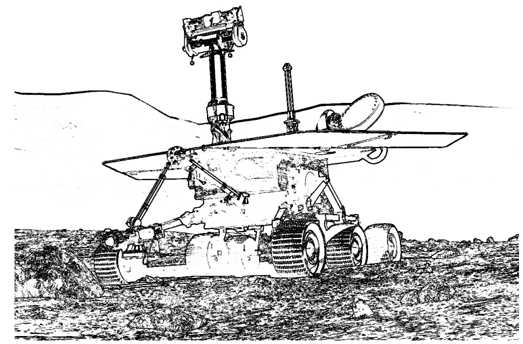




COMP417
Introduction to Robotics and Intelligent Systems

SLAM

Slides from Florian Shkurti,
Andrew W. Moore (CMU), Richard D. Kass (Ohio State)
And other sources

McGill

MRL Mobile Robotics Lab at McGill University

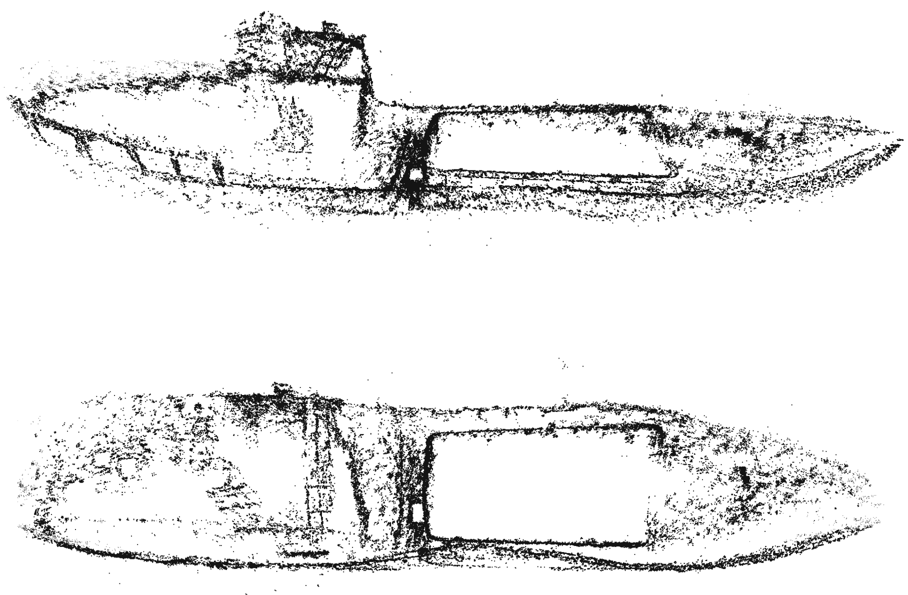
Goal

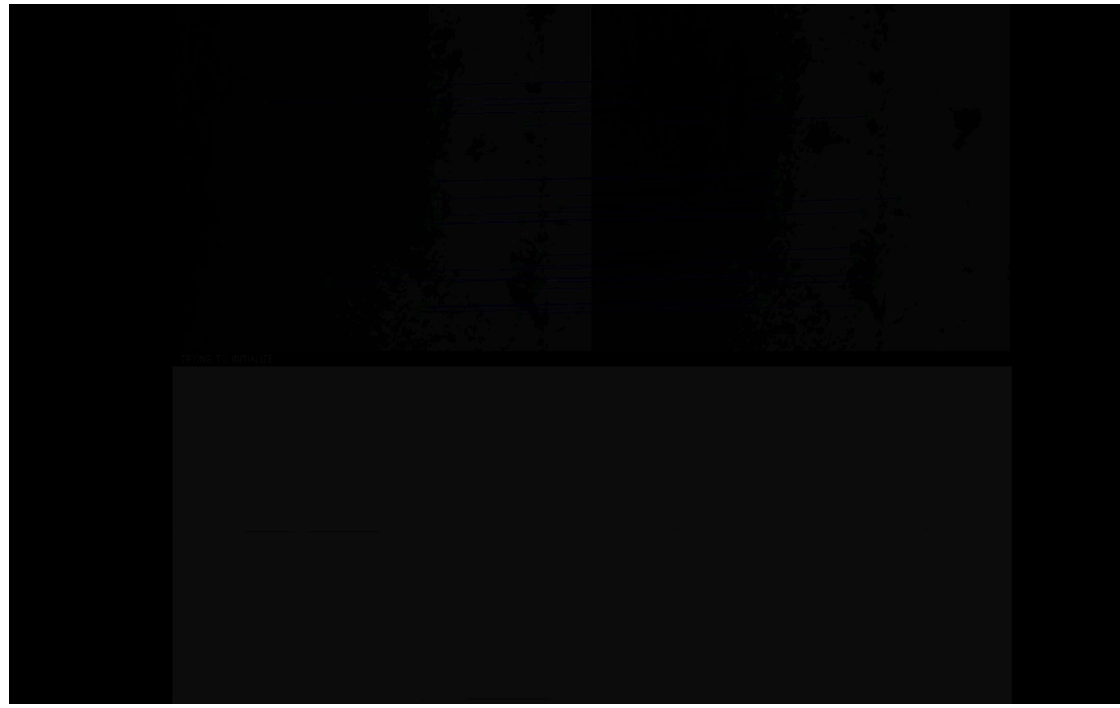
- Enable a robot to simultaneously build a map of its environment and estimate where it is in that map.
- This is called SLAM (Simultaneous Localization And Mapping)

Goal

- Enable a robot to simultaneously build a map of its environment and estimate where it is in that map.
- This is called SLAM (Simultaneous Localization And Mapping)
- Today we are going to look at the batch version, i.e. collect all measurements and controls, and later form an estimate of the states and the map.
- We are going to solve SLAM using least squares

Examples of SLAM systems

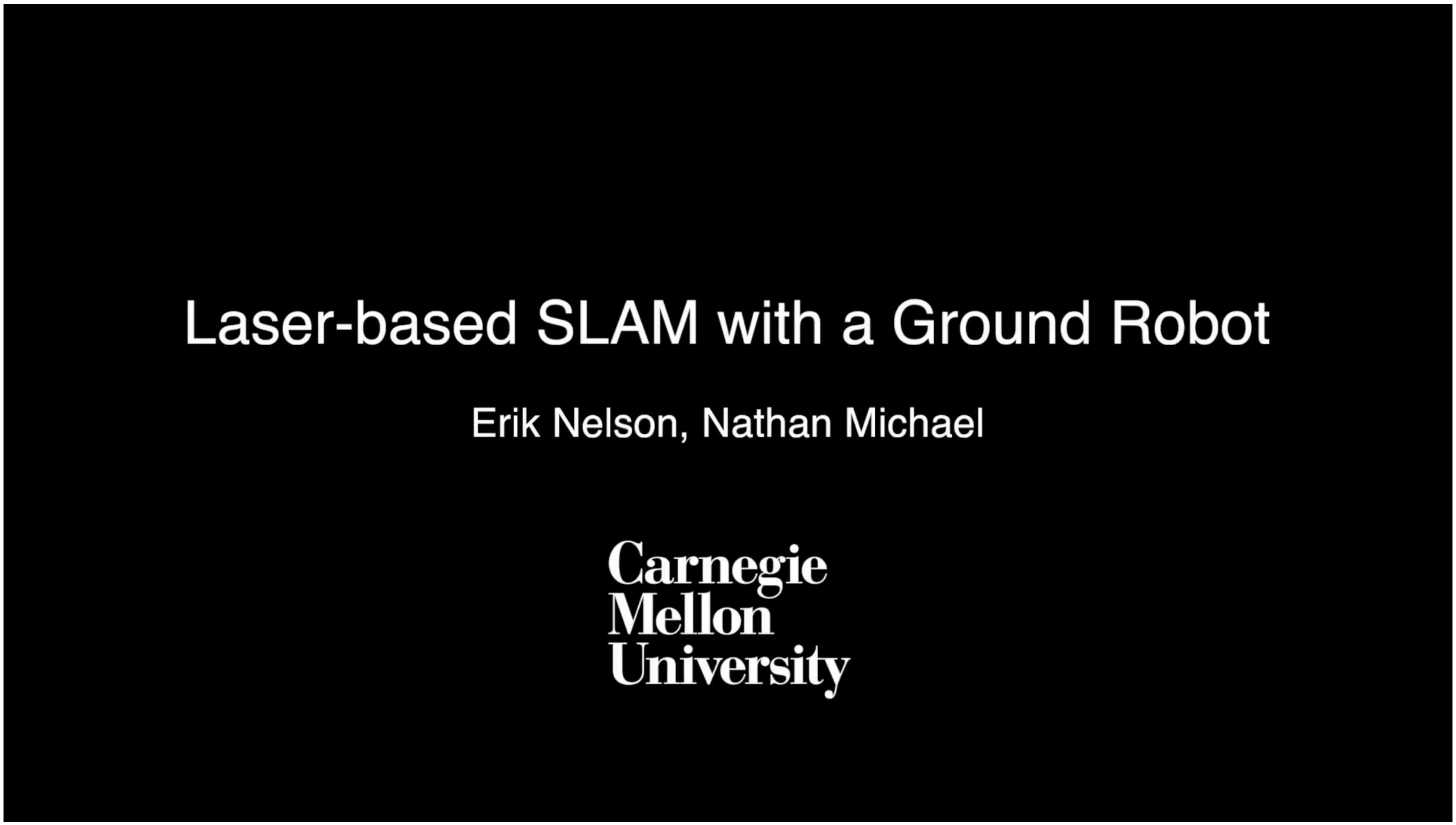




MORESLAM system, McGill, 2016

MORESLAM system, McGill, 2016

Examples of SLAM systems

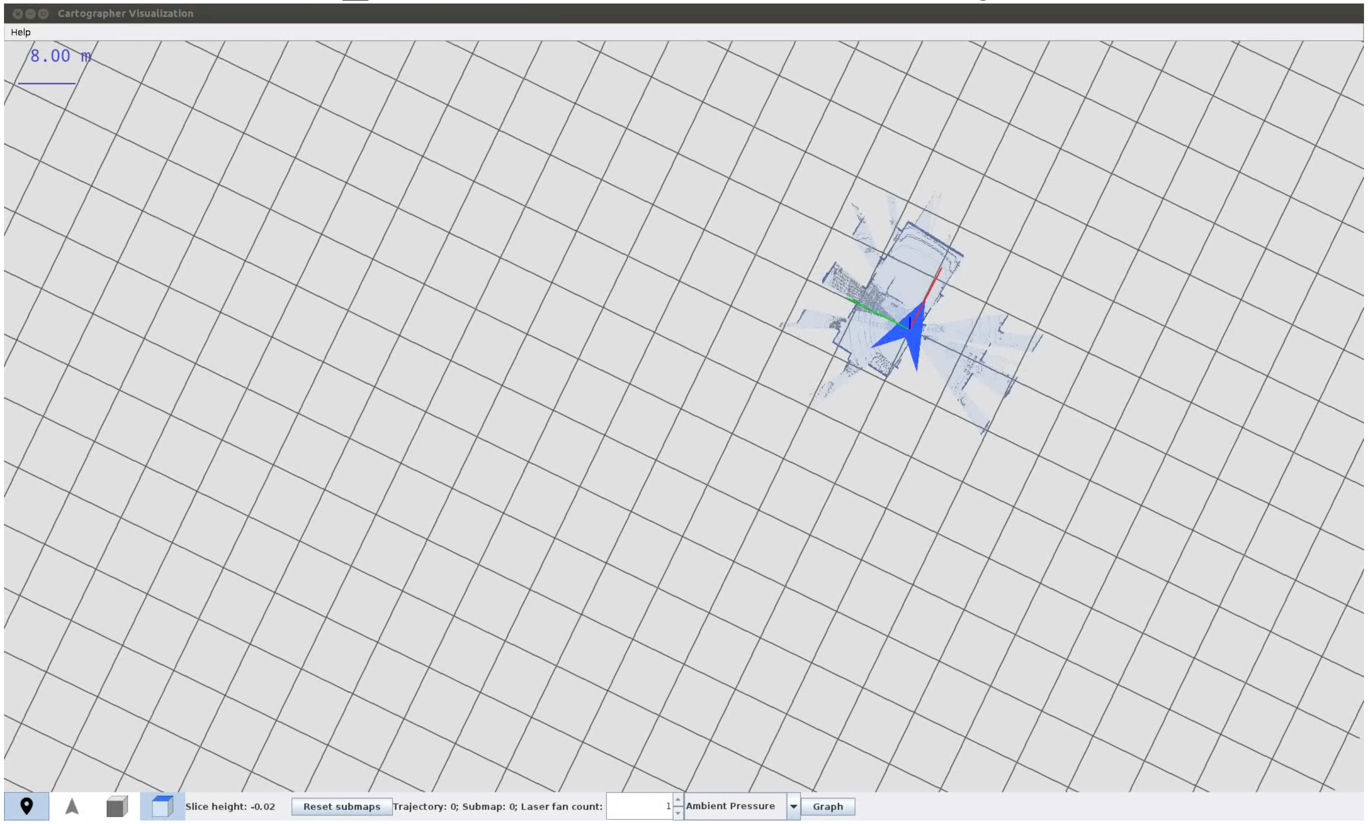


Examples of SLAM systems

Source Code: https://github.com/erik-nelson/blam

Examples of SLAM systems

Google Cartographer: 2D and 3D laser SLAM



Code: https://github.com/googlecartographer/cartographer

SLAM: possible problem definitions

- Smoothing/Batch/Full SLAM

Initial state

$$p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$$

Sequence of states

Map

Sequence of sensor measurements

Sequence of commands

SLAM: possible problem definitions

- Smoothing/Batch/Full SLAM

Initial state

$$p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$$

Sequence of states

Map

Sequence of sensor measurements

Sequence of commands

- Filtering SLAM

$$p(\mathbf{x}_t, \mathbf{m}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}, \mathbf{x}_0)$$

SLAM: possible problem definitions

In this lecture

- Smoothing/Batch/Full SLAM

Initial state

$$p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$$

Sequence of states

Map

Sequence of sensor measurements

Sequence of commands

- Filtering SLAM

$$p(\mathbf{x}_t, \mathbf{m}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}, \mathbf{x}_0)$$

Side topic (backgrounder)
The Gaussian Probability Distribution Function

Introduction

- The Gaussian probability distribution is perhaps the most used distribution in all of science.
  Sometimes it is called the "bell shaped curve" or *normal* distribution.
- Unlike the binomial and Poisson distribution, the Gaussian is a continuous distribution:

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

  μ = mean of distribution (also at the same place as mode and median)
  σ^2 = variance of distribution
  y is a continuous variable ($-\infty \leq y \leq \infty$)

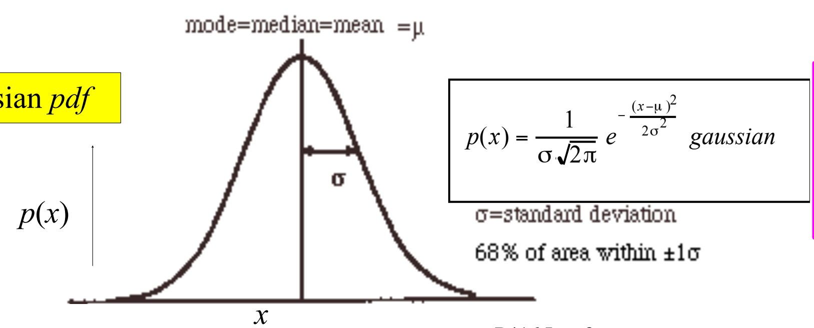
Karl Friedrich Gauss 1777-1855

- Probability (P) of y being in the range $[a, b]$ is given by an integral:

$$P(a < y < b) = \int_a^b p(y)dy = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

  - The integral for arbitrary a and b cannot be evaluated analytically.
    The value of the integral has to be looked up in a table (e.g. Appendixes A and B of Taylor).

Plot of Gaussian *pdf*



The integrals with limits $[-\infty, \infty]$ can be evaluated, see Barlow P. 37.

$$\lim_{n\to\infty} P\left[a < \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} < b\right] = \lim_{n\to\infty} P\left[a < \frac{\bar{Y}-\mu}{\sigma_m} < b\right] = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}\,dy$$

— σ_m is sometimes called "the error in the mean" (more on that later):

$$\sigma_m = \frac{\sigma}{\sqrt{n}}$$

☞ For CLT to be valid:

The μ and σ of the *pdf* must be finite.

No one term in sum should dominate the sum.

- A random variable is not the same as a random number.

  "A random variable is any rule that associates a number with each outcome in S" (Devore, in "probability and Statistics for Engineering and the Sciences").

  Here S is the set of possible outcomes.

- Recall if y is described by a Gaussian *pdf* with mean (μ) of zero and σ=1 then the probability that a<y<b is given by:

$$P(a < y < b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}\,dy$$

The CLT is true even if the Y's are from different *pdf*'s as long as the means and variances are defined for each *pdf* !

See Appendix of Barlow for a proof of the Central Limit Theorem.

- The total area under the curve is normalized to one by the $\sigma\sqrt{(2\pi)}$ factor.

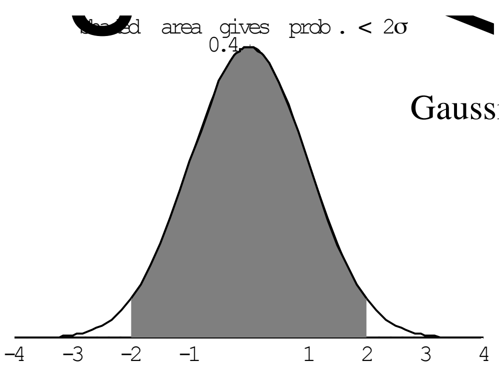
$$P(-\infty < y < \infty) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = 1$$

- We often talk about a measurement being a certain number of standard deviations (σ) away from the mean (μ) of the Gaussian.

We can associate a probability for a measurement to be $|\mu - n\sigma|$ from the mean just by calculating the area outside of this region.

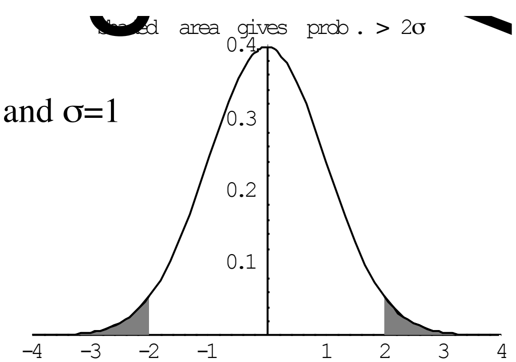
$n\sigma$	Prob. of exceeding $\mu\pm n\sigma$
0.67	0.5
1	0.32
2	0.05
3	0.003
4	0.00006

It is very unlikely (< 0.3%) that a measurement taken at random from a Gaussian *pdf* will be more than ± 3σ from the true mean of the distribution.



95% of area within 2σ

Gaussian with μ=0 and σ=1



Only 5% of area outside 2σ

Why is the Gaussian *pdf* so applicable? ⇒ Central Limit Theorem

A crude statement of the Central Limit Theorem:

Things that are the result of the addition of lots of small effects tend to become Gaussian.

A more exact statement:

Let $Y_1, Y_2, \ldots Y_n$ be an infinite sequence of independent random variables each with the same probability distribution.

Suppose that the mean (μ) and variance (σ^2) of this distribution are both finite.

For any numbers a and b:

$$\lim_{n\to\infty} P\left[a < \frac{Y_1 + Y_2 + \ldots Y_n - n\mu}{\sigma\sqrt{n}} < b\right] = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}\,dy$$

Actually, the Y's can be from different *pdf*'s!

The C.L.T. tells us that under a wide range of circumstances the probability distribution that describes the sum of random variables tends towards a Gaussian distribution as the number of terms in the sum $\to\infty$.

How close to ∞ does n have to be??

Alternatively we can write the CLT in a different form:

$$\lim_{n\to\infty} P\left[a < \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < b\right] = \lim_{n\to\infty} P\left[a < \frac{\bar{Y} - \mu}{\sigma_m} < b\right] = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}\,dy$$

Expectation (mean)

- Expected value of a random variable X:

$$E_{p(X)}[X] = \int_x x p(X = x) dx$$

Probability-weighted average

- E is linear:

$$E_{p(X)}[X + c] = E_{p(X)}[X] + c$$

$$E_{p(X)}[AX + b] = AE_{p(X)}[X] + b$$

If every student's grade goes up by 10, the class average goes up by 10

- If X,Y are independent then [Note: inverse does not hold]

$$E_{p(X,Y)}[XY] = E_{p(X)}[X]E_{p(Y)}[Y]$$

Entropy of a PDF (looks a bit like expectation)

$$\text{Entropy of } X = H[X] = -\int_{x=-\infty}^{\infty} p(x) \log p(x) dx$$

Natural log (ln or $\log_e$)

The larger the entropy of a distribution…

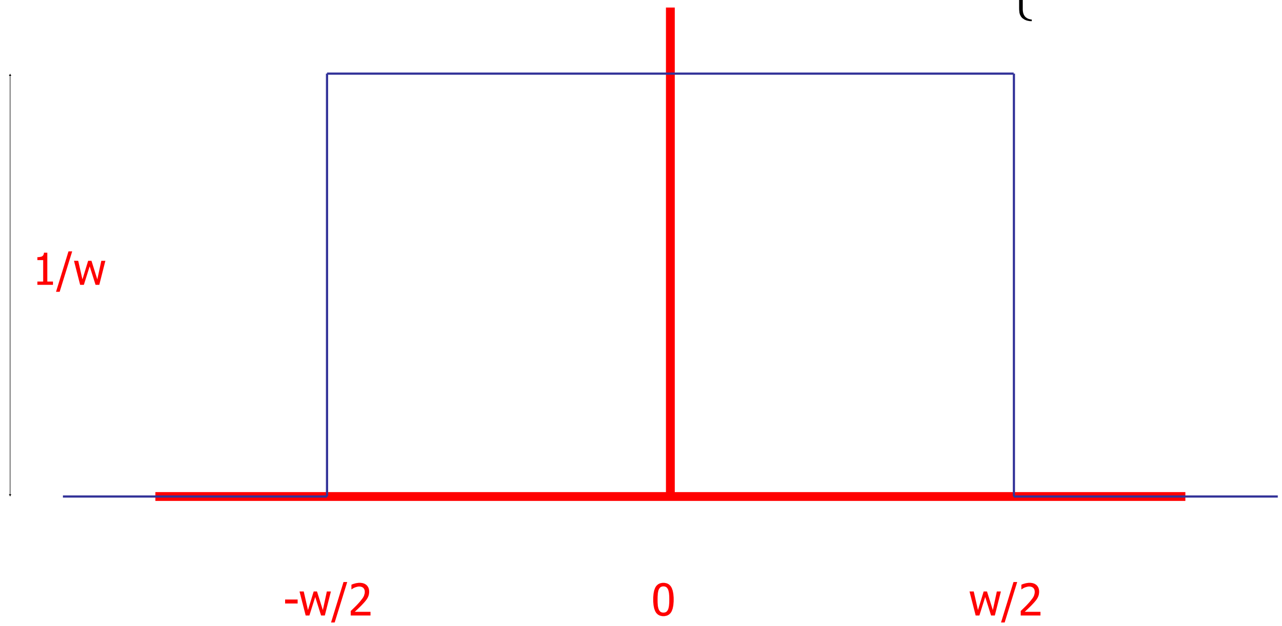
…the harder it is to predict

…the harder it is to compress it

…the less spiky the distribution

The "box" distribution

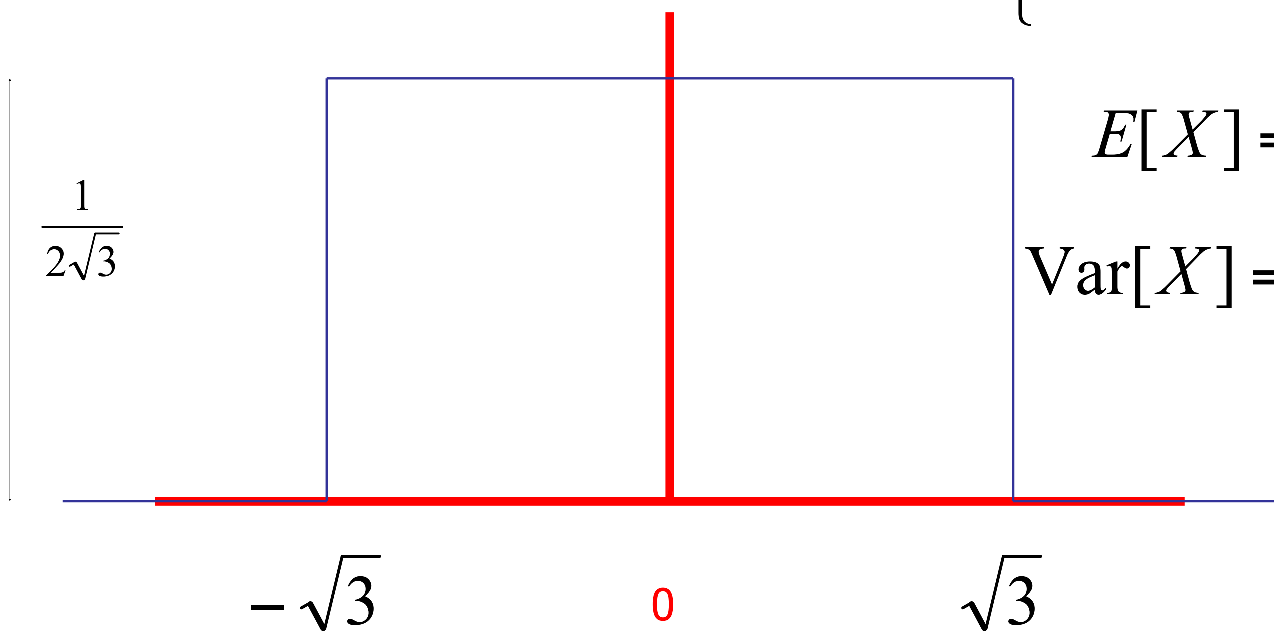
$$p(x) = \begin{cases} \frac{1}{w} & \text{if} \quad |x| \le \frac{w}{2} \\ 0 & \text{if} \quad |x| > \frac{w}{2} \end{cases}$$



$$H[X] = -\int_{x=-\infty}^{\infty} p(x) \log p(x) dx = -\int_{x=-w/2}^{w/2} \frac{1}{w} \log \frac{1}{w} dx = -\frac{1}{w} \log \frac{1}{w} \int_{x=-w/2}^{w/2} dx = \log w$$

Unit variance box distribution

$$p(x) = \begin{cases} \frac{1}{w} & \text{if} \quad |x| \le \frac{w}{2} \\ 0 & \text{if} \quad |x| > \frac{w}{2} \end{cases}$$



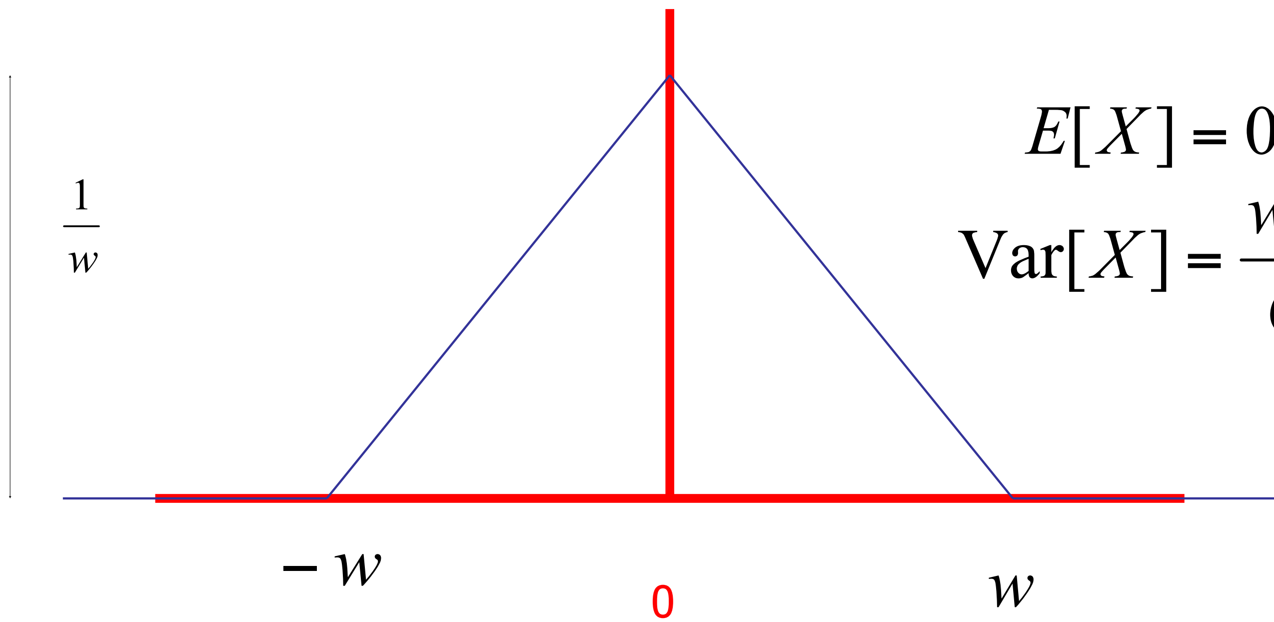
$$E[X] = 0$$

$$\mathrm{Var}[X] = \frac{w^2}{12}$$

$$\text{if } w = 2\sqrt{3} \text{ then } \mathrm{Var}[X] = 1 \text{ and } H[X] = 1.242$$

The Hat distribution

$$p(x) = \begin{cases} \dfrac{w - |x|}{w^2} & \text{if} \quad |x| \le w \\ 0 & \text{if} \quad |x| > w \end{cases}$$

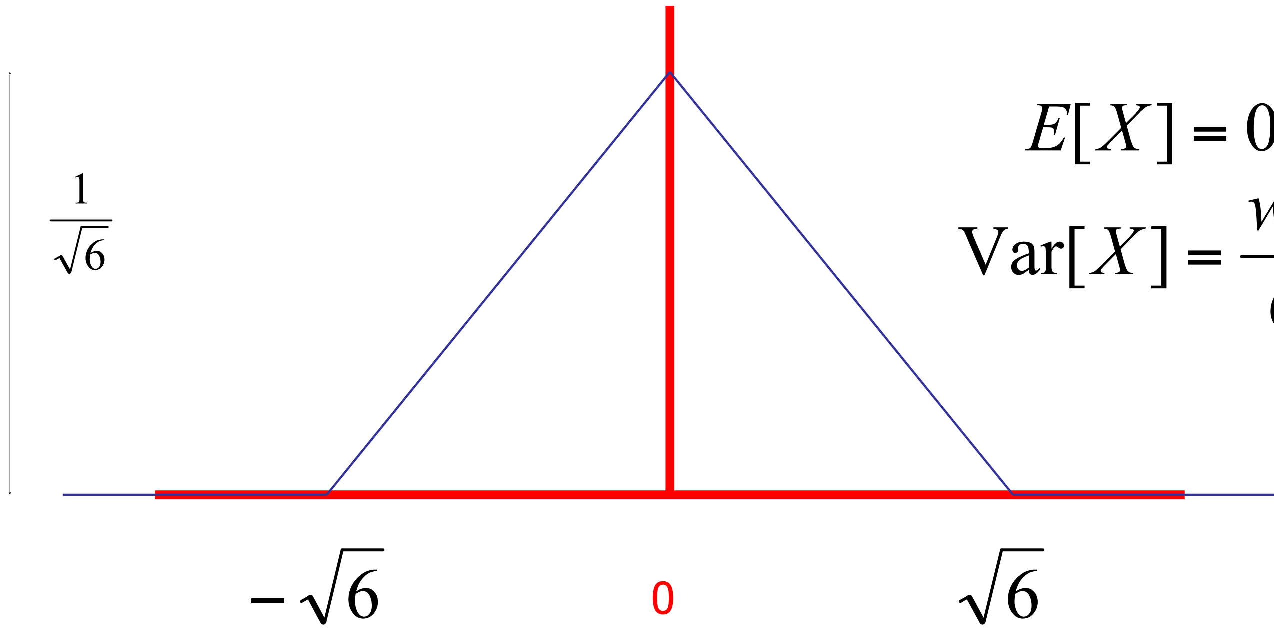


$$E[X] = 0$$

$$\mathrm{Var}[X] = \frac{w^2}{6}$$

Unit variance hat distribution

$$p(x) = \begin{cases} \dfrac{w - |x|}{w^2} & \text{if} \quad |x| \le w \\ 0 & \text{if} \quad |x| > w \end{cases}$$



$$E[X] = 0$$

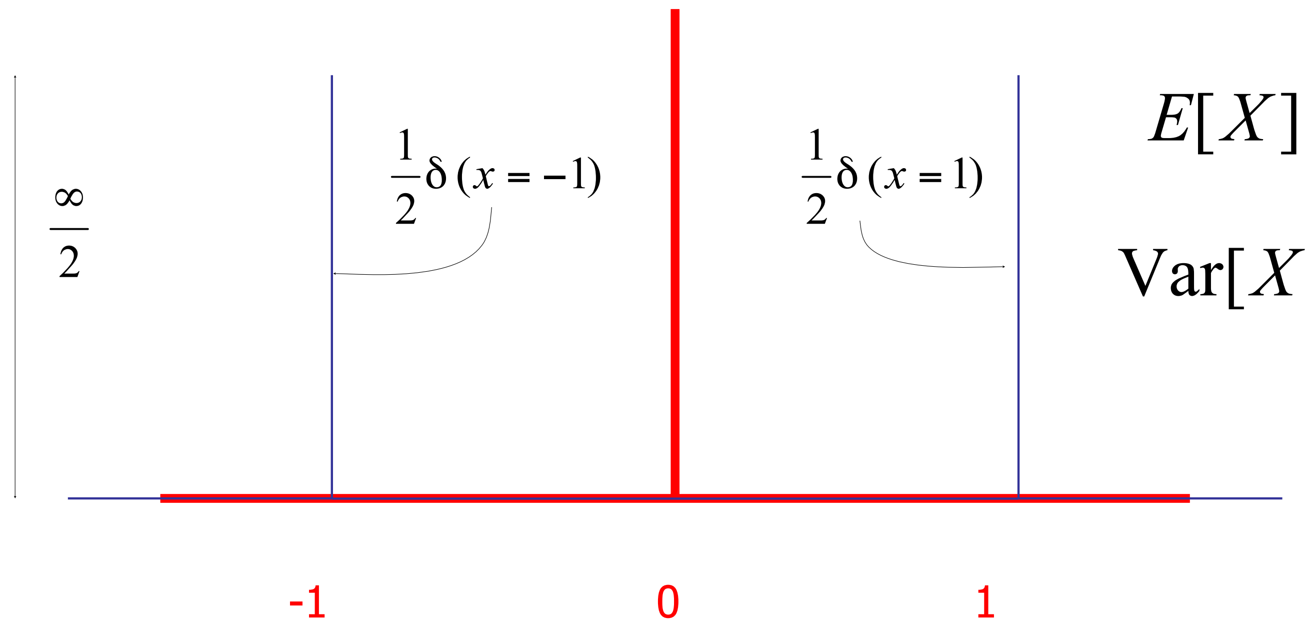
$$\mathrm{Var}[X] = \frac{w^2}{6}$$

$$\text{if } w = \sqrt{6} \text{ then } \mathrm{Var}[X] = 1 \text{ and } H[X] = 1.396$$

The “2 spikes” distribution

Dirac Delta

$$p(x) = \frac{\delta(x=-1) + \delta(x=1)}{2}$$



$$E[X] = 0$$

$$\mathrm{Var}[X] = 1$$

$$H[X] = -\int_{x=-\infty}^{\infty} p(x) \log p(x) dx = -\infty$$

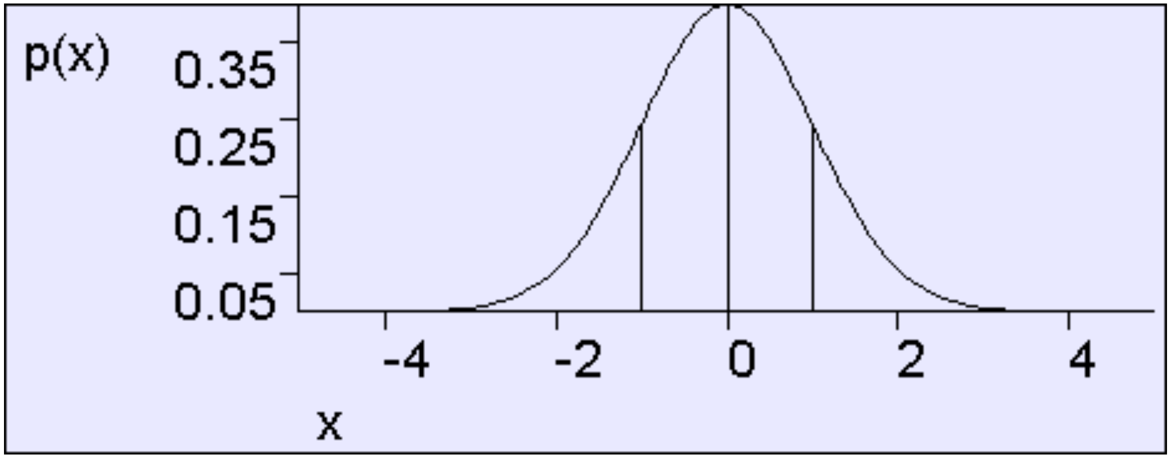
Entropies of unit-variance distributions

Distribution	Entropy
Box	1.242
Hat	1.396
2 spikes	-infinity
???	1.4189

Largest possible entropy of any unit-variance distribution

Unit variance Gaussian

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



$$E[X] = 0$$

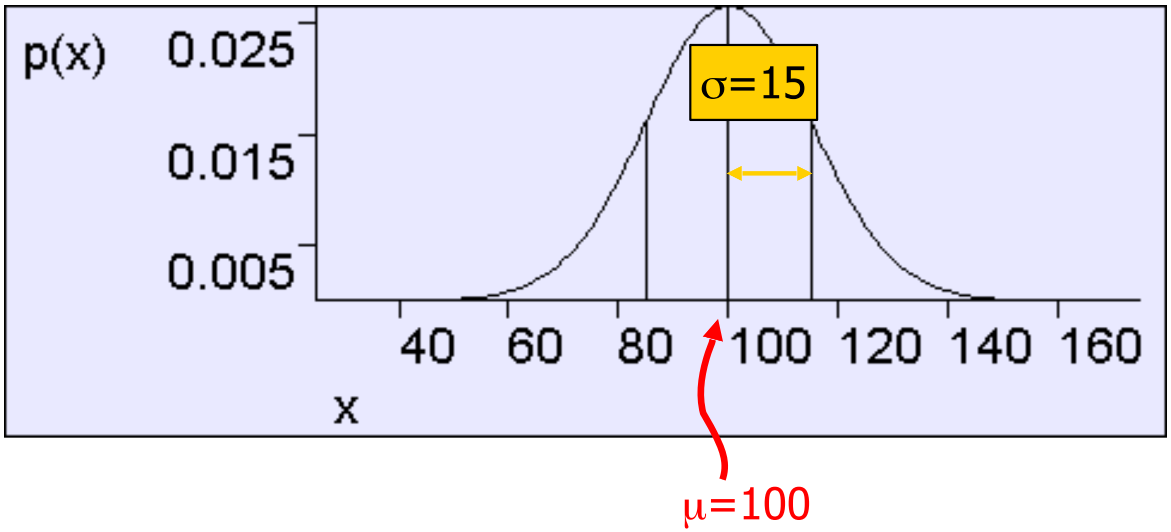
$$\mathrm{Var}[X] = 1$$

$$H[X] = -\int_{x=-\infty}^{\infty} p(x) \log p(x) dx = 1.4189$$

General Gaussian

Also known as the normal distribution or Bell-shaped curve

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



$$E[X] = \mu$$

$$\mathrm{Var}[X] = \sigma^2$$

Shorthand: We say X ~ N(μ,σ^2) to mean "X is distributed as a Gaussian with parameters μ and σ^2".

In the above figure, X ~ N(100,15^2)

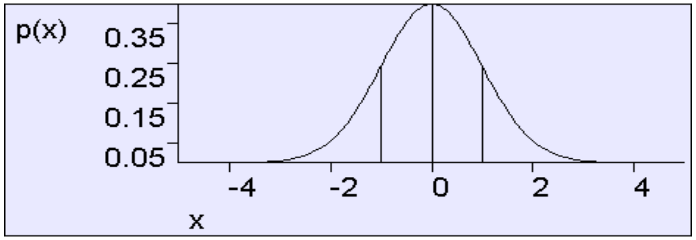
The Error Function

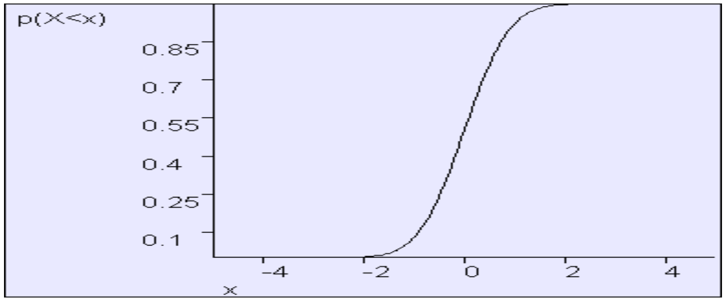
Assume X ~ N(0,1)

Define ERF(x) = P(X<x) = Cumulative Distribution of X

$$ERF(x) = \int_{z=-\infty}^{x} p(z)dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{z=-\infty}^{x} \exp\left(-\frac{z^2}{2}\right)dz$$

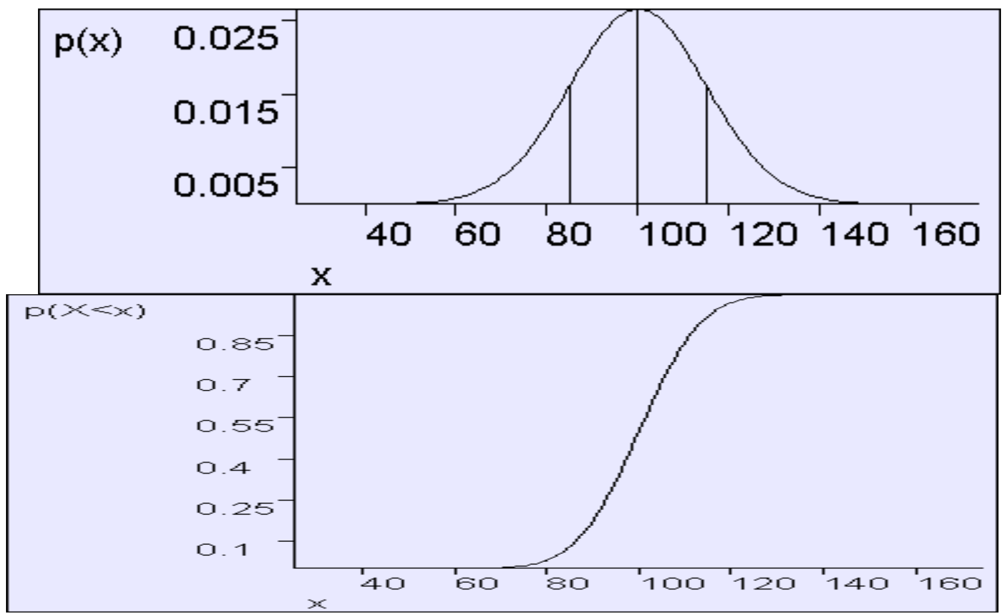




Using The Error Function

Assume X ~ N(μ,σ^2)

P(X<x| μ,σ^2) = $ERF(\frac{x-\mu}{\sigma^2})$



Covariance Matrix

- How two (or more) variables X and Y (random variables) are related. Measures linear dependence between random variables X, Y.
  Does **not** measure independence.

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

- Variance of X

Terminology:
***variance** of*
*a random multi-dimensional variable is also known as **covariance** matrix.*

$$\text{Var}[X] = \text{Cov}[X] = \text{Cov}[X, X] = E[X^2] - E[X]^2$$

$$\text{Cov}[AX + b] = A\text{Cov}[X]A^T$$

$$\text{Cov}[X + Y] = \text{Cov}[X] + \text{Cov}[Y] - 2\text{Cov}[X, Y]$$

Covariance Matrix

- Measures linear dependence between random variables X, Y. Does **not** measure independence.

$$\mathrm{Cov}[X,Y] = E[XY] - E[X]E[Y]$$

- Entry (i,j) of the covariance matrix measures whether changes in variable X_i co-occur with changes in variable Y_j

- It does not measure whether one causes the other.

Bivariate Gaussians

Write r.v. $\mathbf{X} = \begin{pmatrix} X \\ Y \end{pmatrix}$ Then define $X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to mean

$$p(\mathbf{x}) = \frac{1}{2\pi \, \| \boldsymbol{\Sigma} \|^{1/2}} \exp\left(-\tfrac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \, \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$$

Where the Gaussian's parameters are…

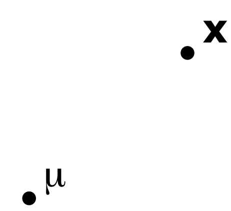
$$\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2{}_x & \sigma_{xy} \\ \sigma_{xy} & \sigma^2{}_y \end{pmatrix}$$

Where we insist that Σ is symmetric non-negative definite

Evaluating p(**x**): Step 1

$$p(\mathbf{x}) = \frac{1}{2\pi \,||\, \boldsymbol{\Sigma} \,||^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \,\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

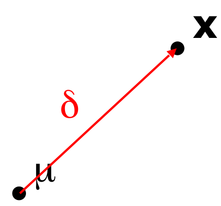
1. Begin with vector **x**



Evaluating p(**x**): Step 2

$$p(\mathbf{x}) = \frac{1}{2\pi \, \|\, \boldsymbol{\Sigma} \,\|^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \, \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

1. Begin with vector **x**
2. Define δ **= x -** μ

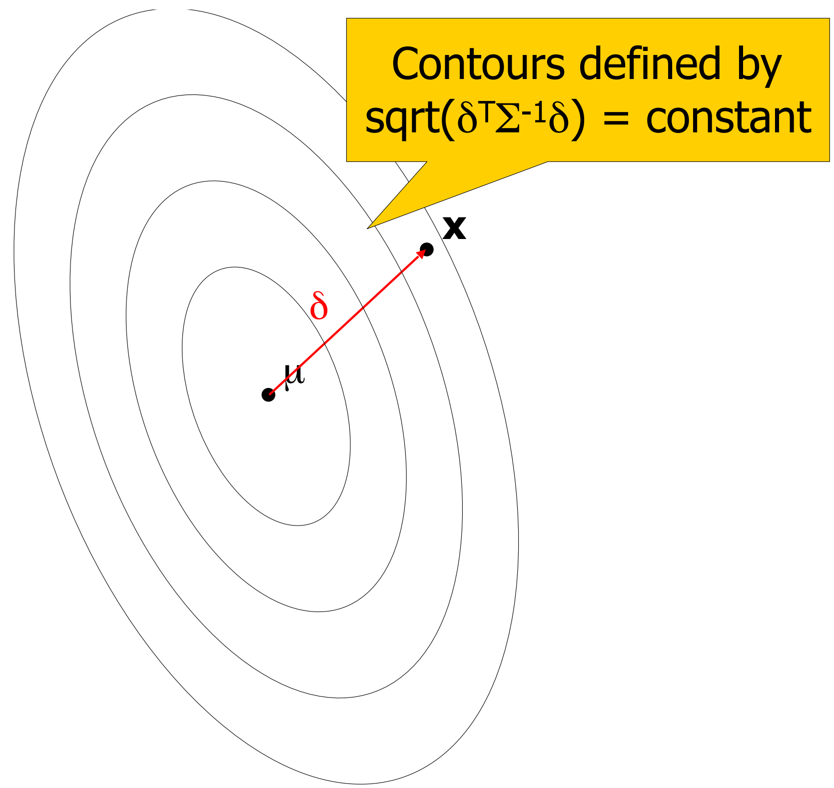


Evaluating p(**x**): Step 3

$$p(\mathbf{x}) = \frac{1}{2\pi \, \|\, \boldsymbol{\Sigma} \,\|^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \, \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

1. Begin with vector **x**
2. Define δ **= x -** μ
3. Count the number of contours crossed of the ellipsoids formed Σ^{-1}

   D = this count = sqrt($\delta^T\Sigma^{-1}\delta$) = Mahalonobis Distance between **x** and μ

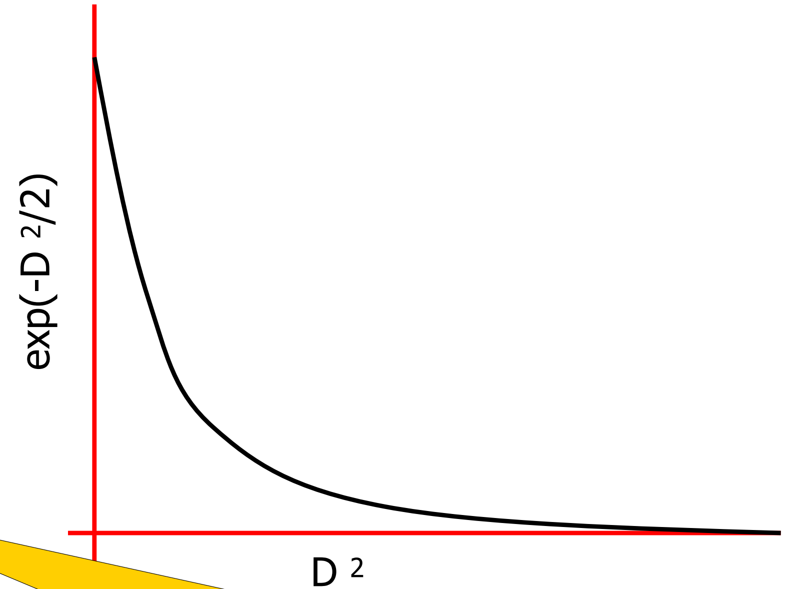


Evaluating p($\mathbf{x}$): Step 4

$$p(\mathbf{x}) = \frac{1}{2\pi \, || \boldsymbol{\Sigma} ||^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

1. Begin with vector **x**
2. Define δ **= x -** μ
3. Count the number of contours crossed of the ellipsoids formed Σ^{-1}

   D = this count = sqrt($\delta^T\Sigma^{-1}\delta$) = Mahalonobis Distance between **x** and μ
4. Define w = exp(-D ^{2}/2)



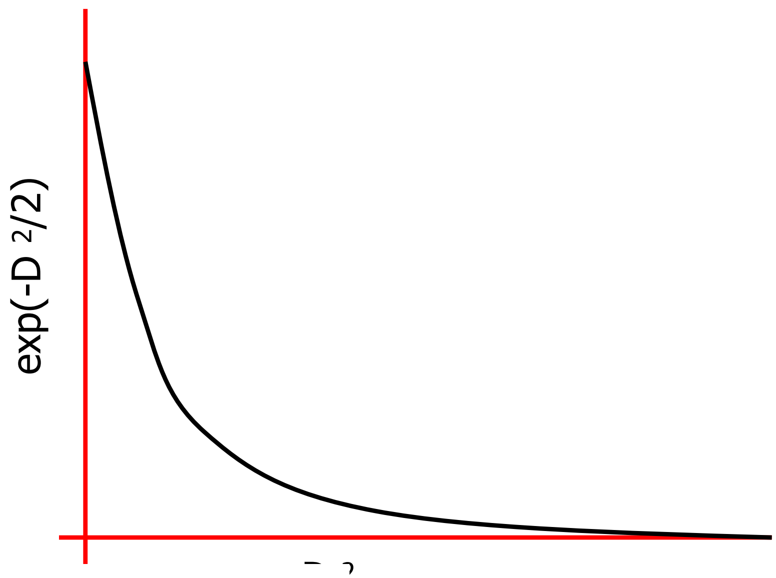
x close to μ in squared Mahalonobis space gets a large weight. Far away gets a tiny weight

Evaluating p(**x**): Step 5

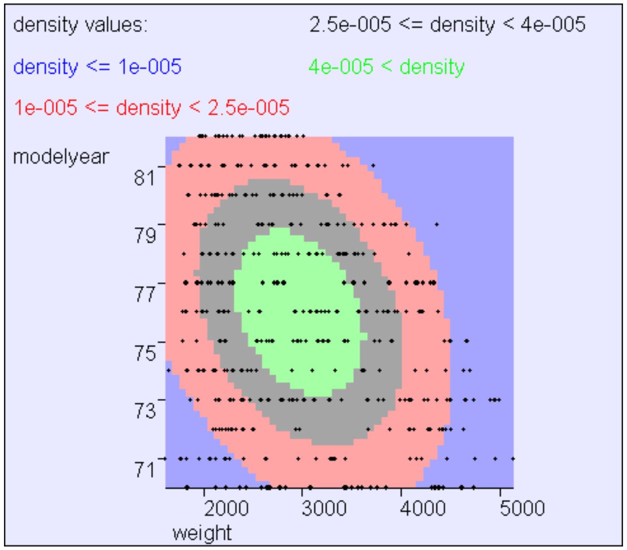
$$p(\mathbf{x}) = \frac{1}{2\pi \, || \boldsymbol{\Sigma} ||^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \, \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

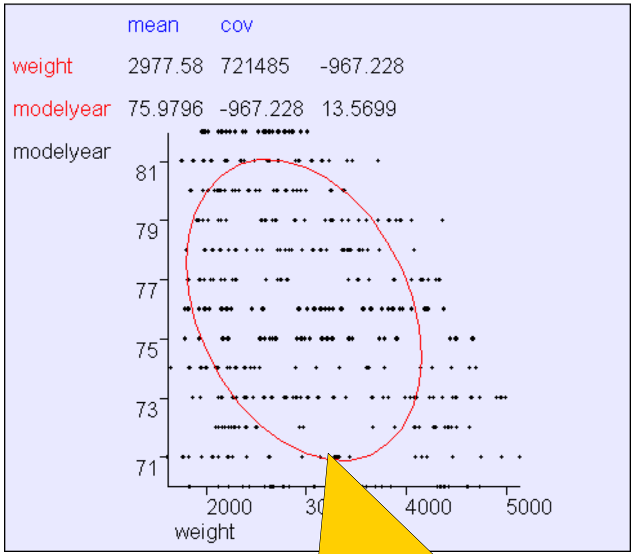
1. Begin with vector **x**
2. Define δ **= x -** μ
3. Count the number of contours crossed of the ellipsoids formed Σ^{-1}

   D = this count = sqrt($\delta^T\Sigma^{-1}\delta$) = Mahalonobis Distance between **x** and μ
4. Define w = exp(-D ^{2}/2)
5. $$\text{Multiply w by } \frac{1}{\sqrt{2\pi} \, || \boldsymbol{\Sigma} ||^{1/2}} \text{ to ensure} \int p(\mathbf{x}) d\mathbf{x} = 1$$



Example

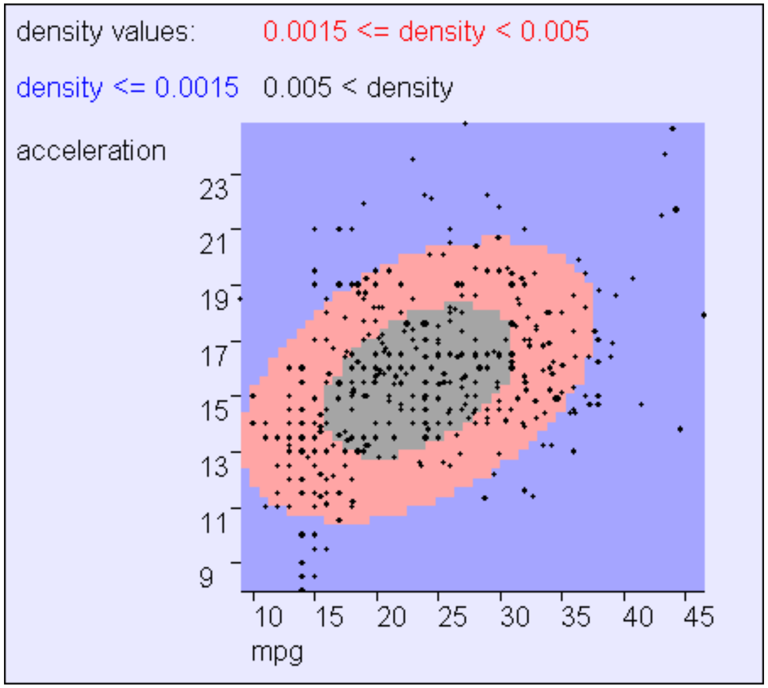


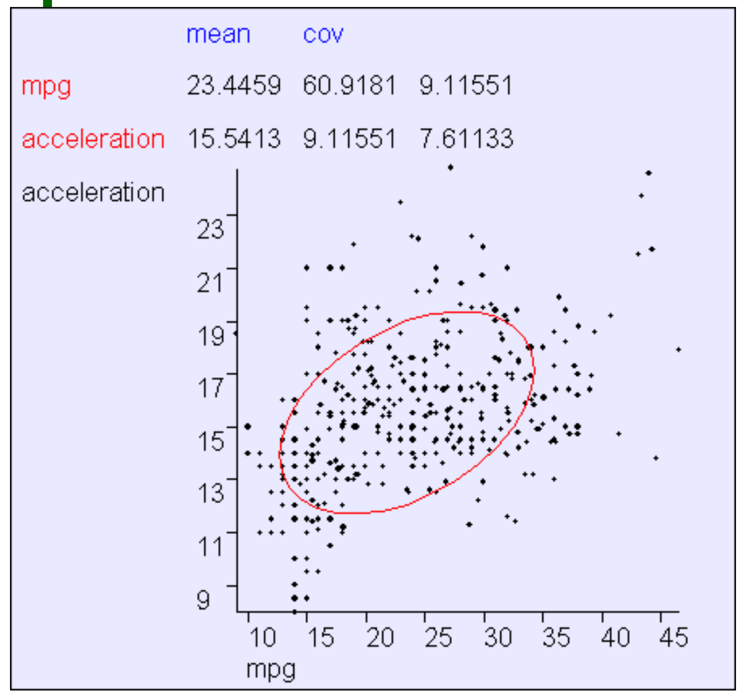


Observe: Mean, Principal axes, implication of off-diagonal covariance term, max gradient zone of p(x)

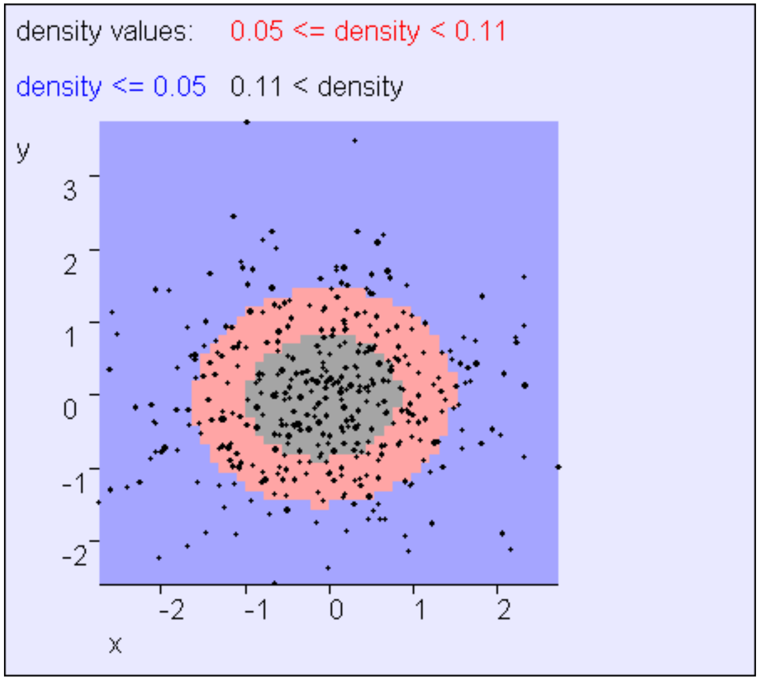
Common convention: show contour corresponding to 2 standard deviations from mean

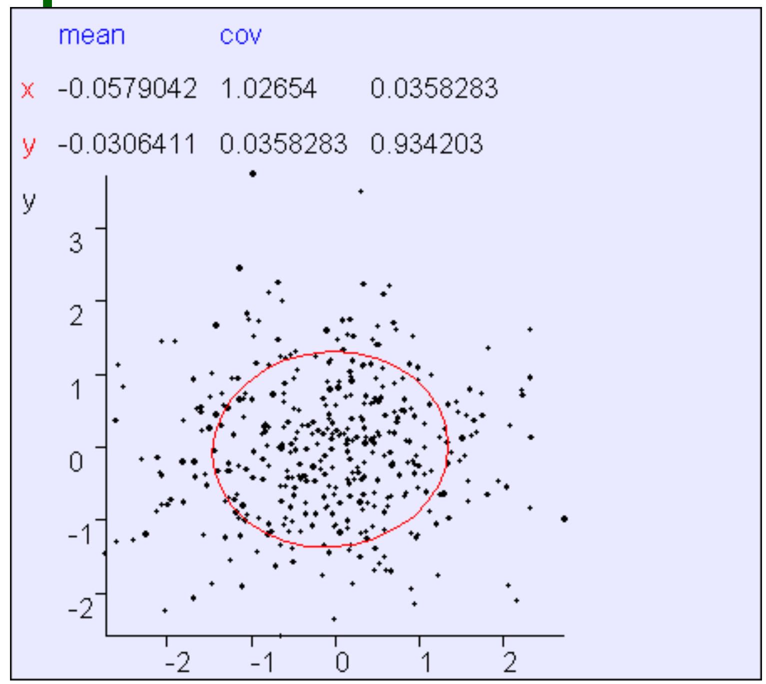
Example


density values:
0.0015 <= density < 0.005
density <= 0.0015
0.005 < density
acceleration
23
21
19
17
15
13
11
9
10 15 20 25 30 35 40 45
mpg


mean
cov
mpg 23.4459 60.9181 9.11551
acceleration 15.5413 9.11551 7.61133
acceleration
23
21
19
17
15
13
11
9
10 15 20 25 30 35 40 45
mpg

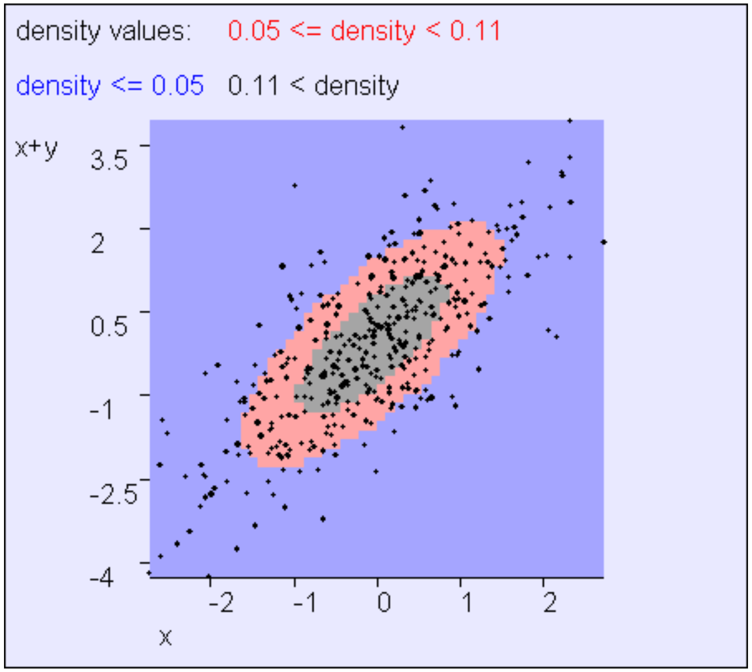
Example

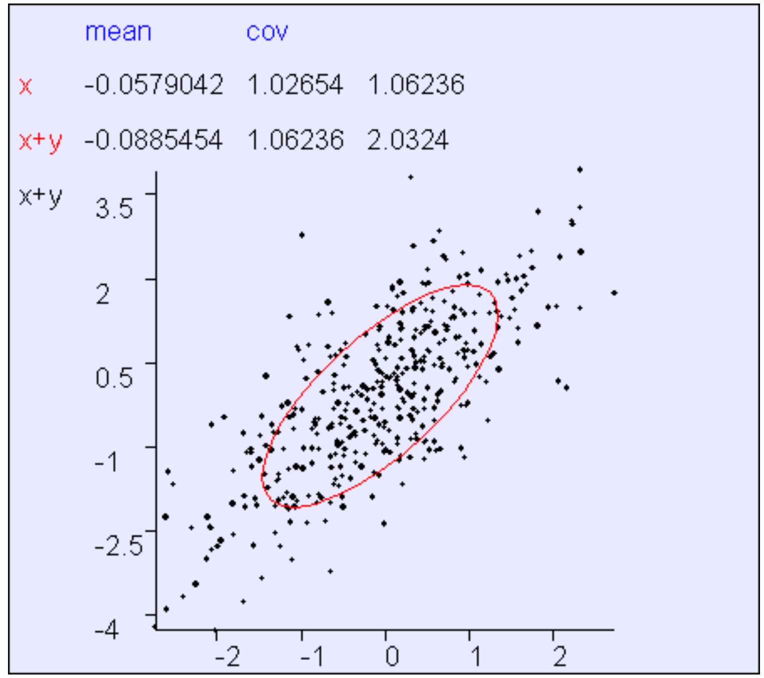




In this example, x and y are almost independent

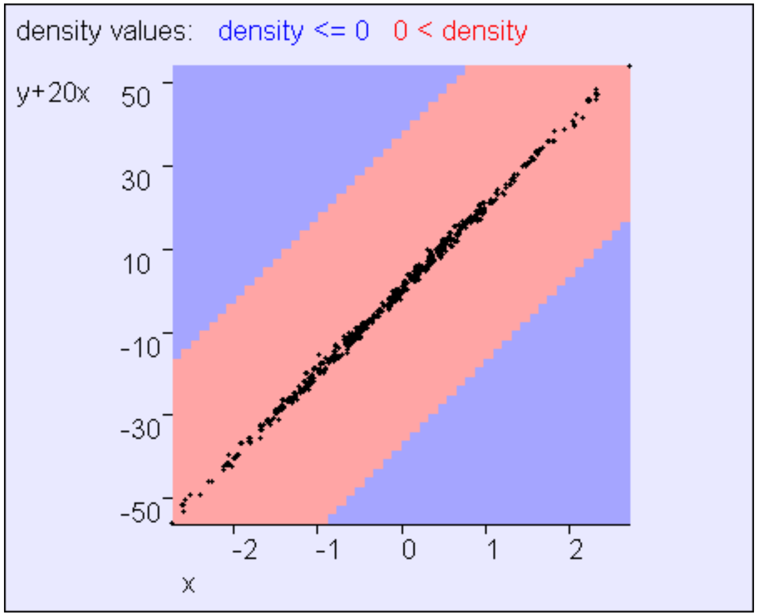
Example

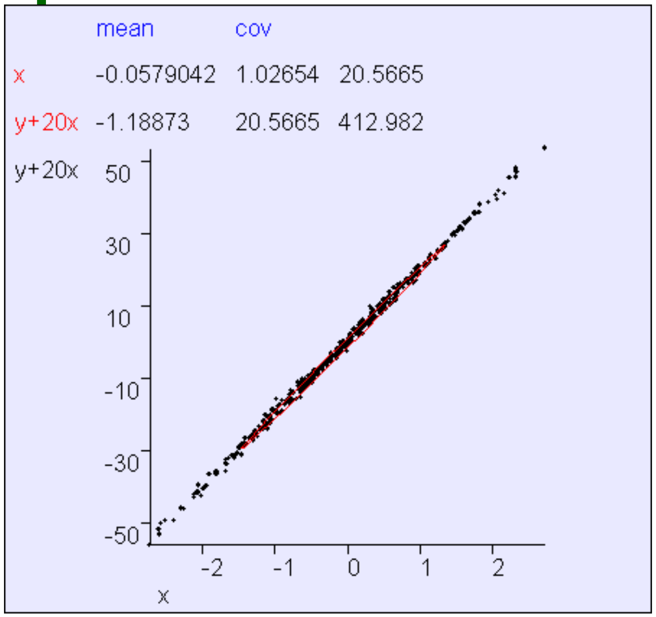




In this example, x and "x+y" are clearly not independent

Example

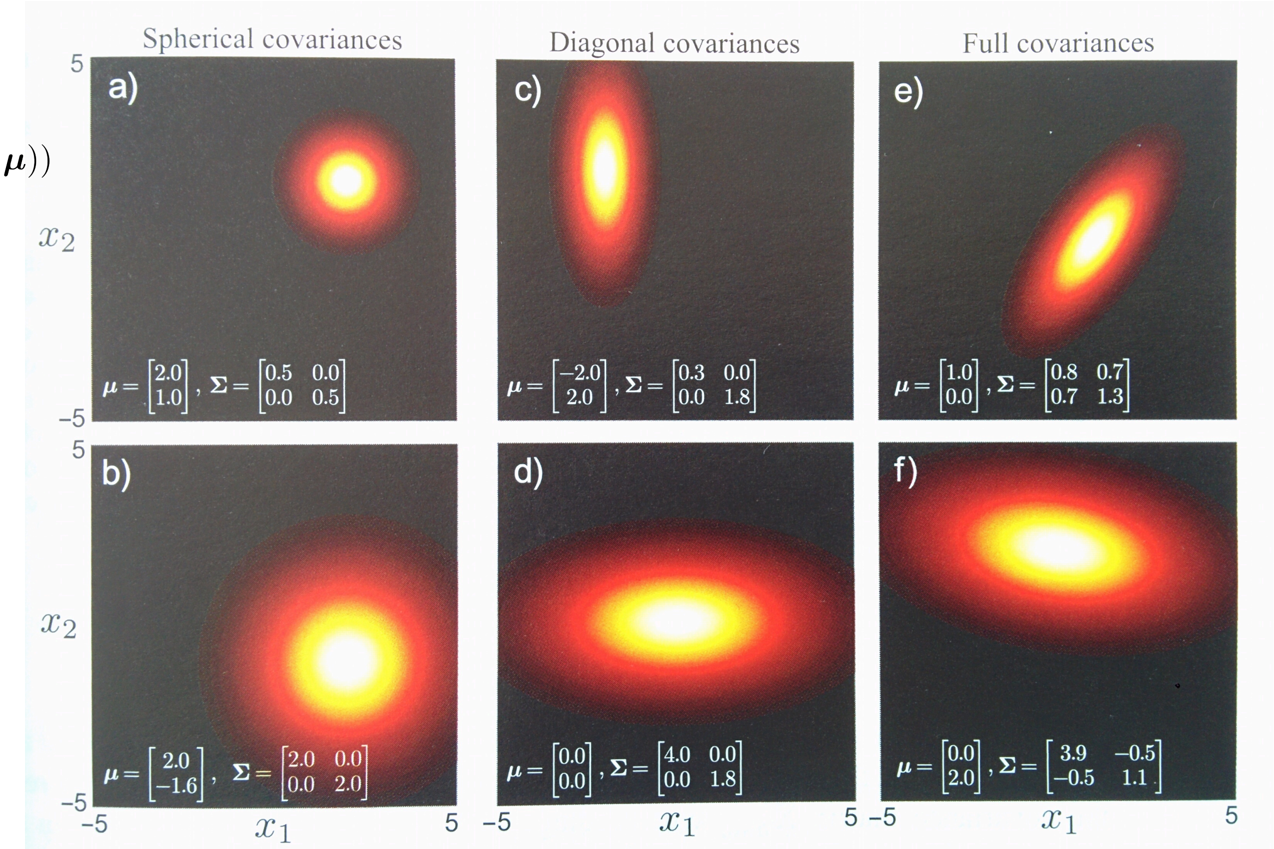




In this example, x and "20x+y" are clearly not independent

Background: Multivariate Gaussian Distribution

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\det(\mathbf{\Sigma})^{1/2}}\exp(-0.5(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))$$



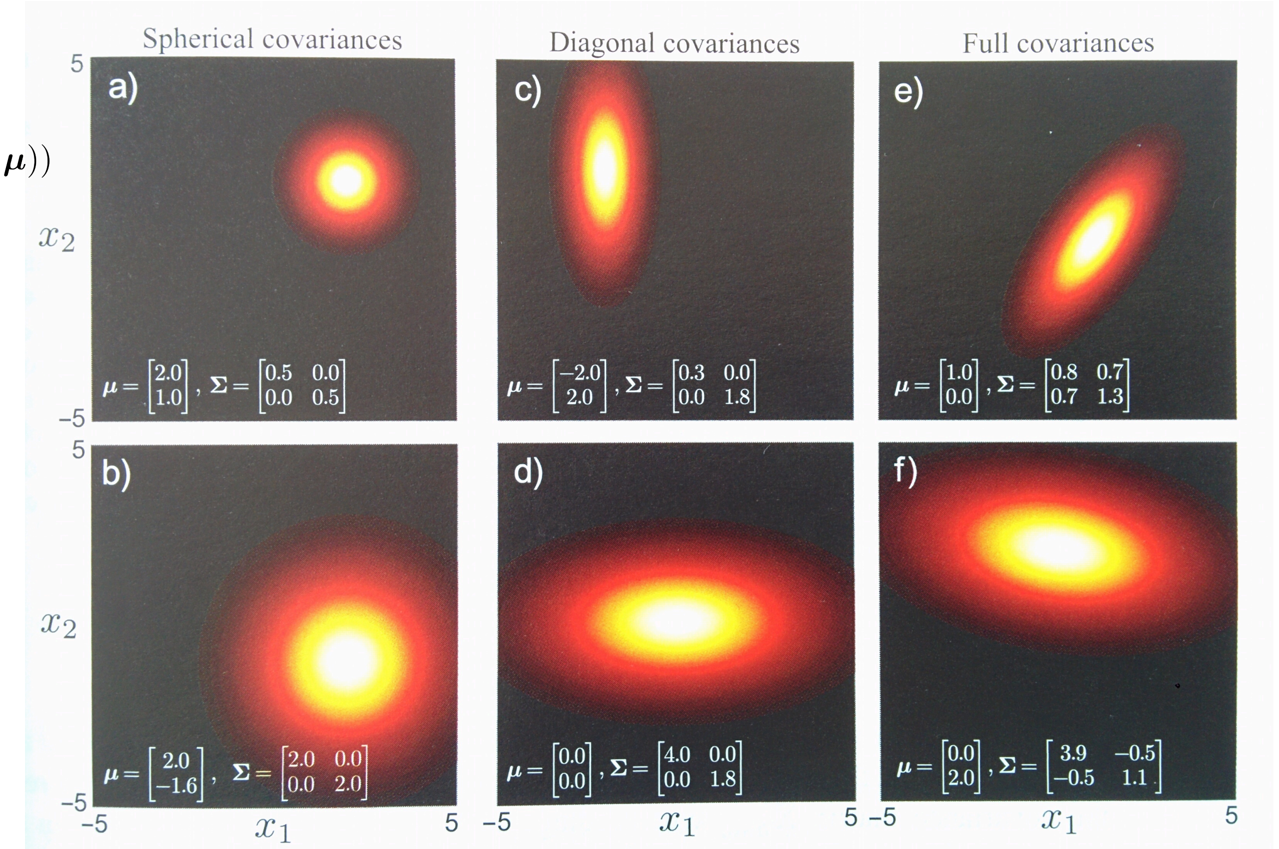
From “Computer Vision: Models, Learning, and Inference” Simon Prince

Background: Multivariate Gaussian Distribution

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\det(\mathbf{\Sigma})^{1/2}}\exp(-0.5(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\det(\mathbf{\Sigma})^{1/2}}\exp(-0.5||\mathbf{x}-\boldsymbol{\mu}||^2_{\mathbf{\Sigma}})$$

Shortcut notation: $||\mathbf{x}||^2_{\mathbf{\Sigma}} = \mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}$



Note: The shapes of these covariances are important, you should know them well. In particular, when are x1 and x2 correlated?

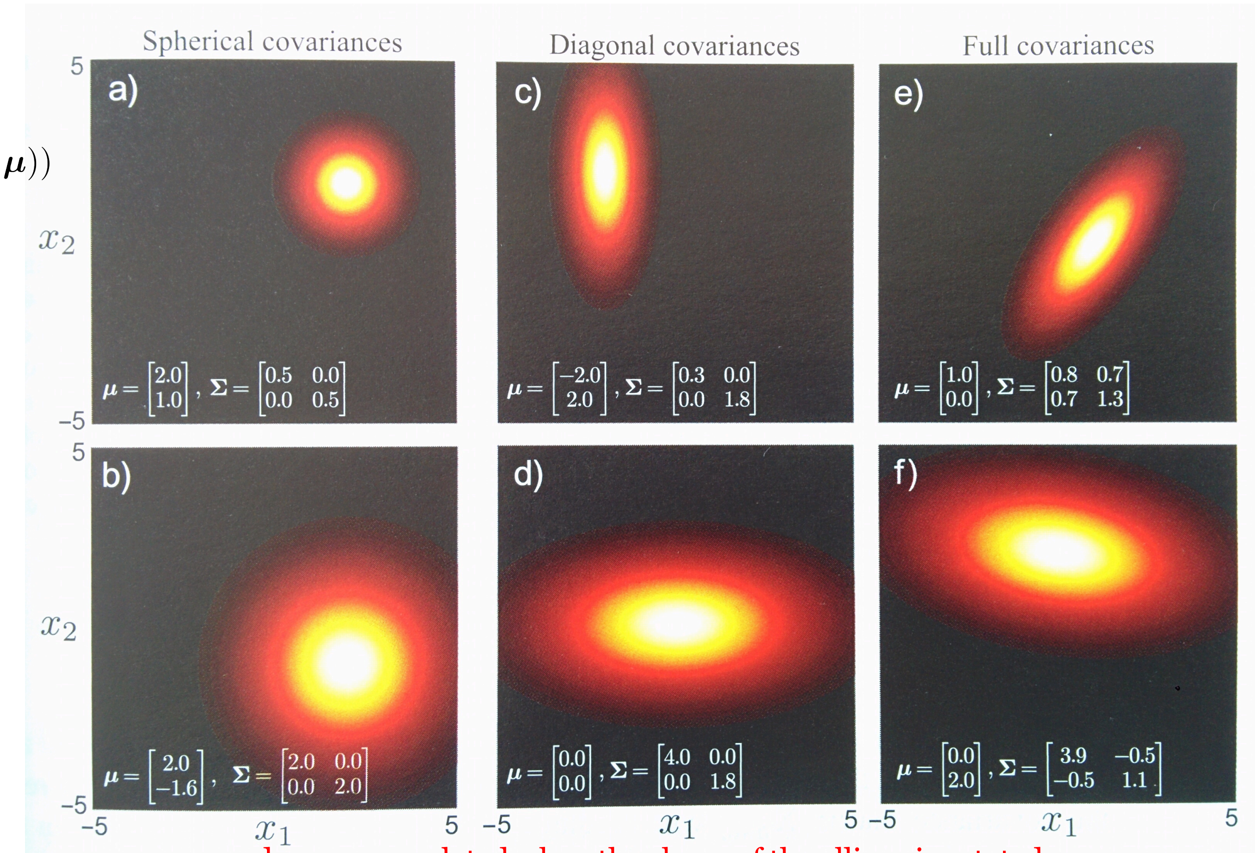
From "Computer Vision: Models, Learning, and Inference" Simon Prince

Background: Multivariate Gaussian Distribution

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\det(\boldsymbol{\Sigma})^{1/2}}\exp(-0.5(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\det(\boldsymbol{\Sigma})^{1/2}}\exp(-0.5||\mathbf{x}-\boldsymbol{\mu}||^2_{\boldsymbol{\Sigma}})$$

Shortcut notation: $||\mathbf{x}||^2_{\boldsymbol{\Sigma}} = \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}$



x1 and x2 are correlated when the shape of the ellipse is rotated,
i.e. when there are nonzero off-diagonal terms in the covariance matrix.
In this example, (e) and (f)

From "Computer Vision: Models, Learning, and Inference" Simon Prince

Confidence regions

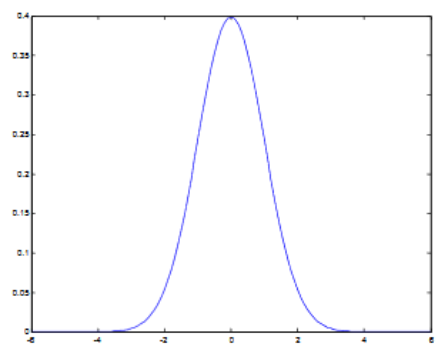
• To quantify confidence and uncertainty define a confidence region R about a point x (e.g. the mode) such that at a confidence level $c \leq 1$

$$p(x \in R) = c$$

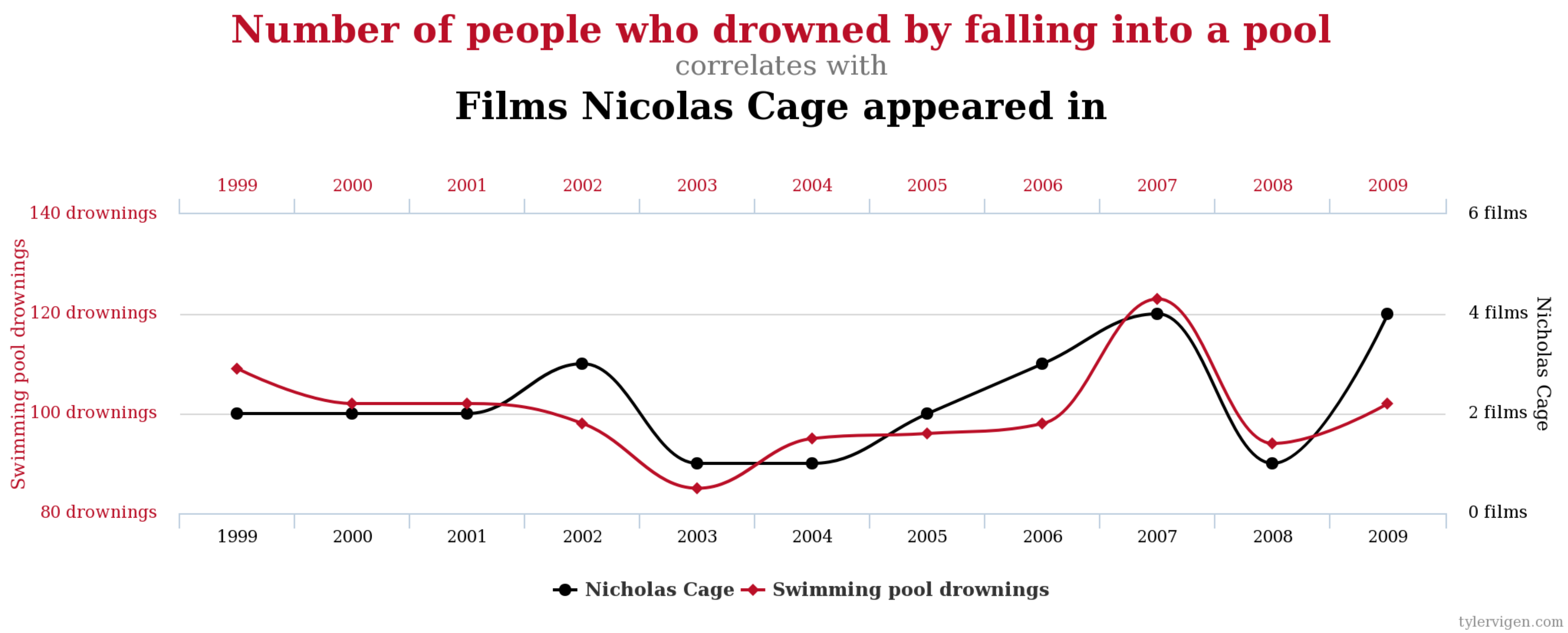
• we can then say (for example) there is a 99% probability that the true value is in R

• e.g. for a univariate normal distribution $N(\mu,\sigma^2)$

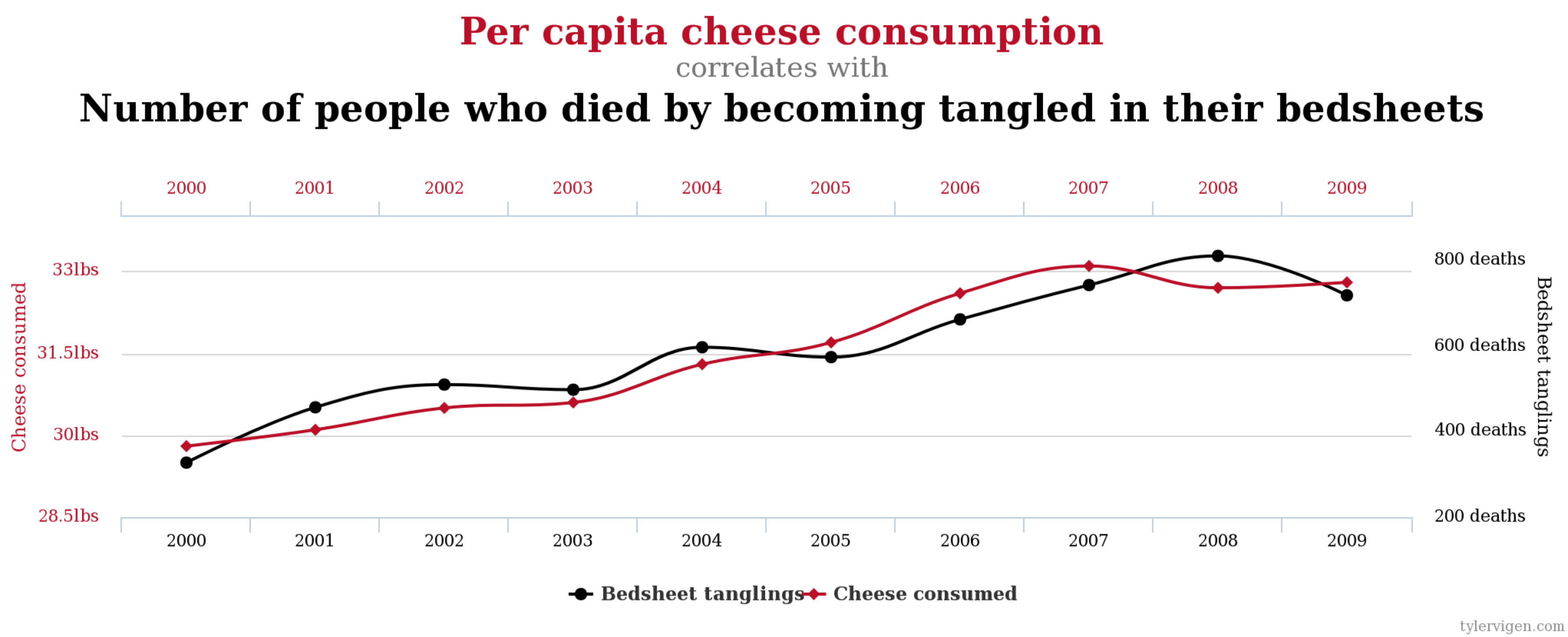
$$\begin{aligned}
p(|x-\mu| < \sigma) &\approx 0.67 \\
p(|x-\mu| < 2\sigma) &\approx 0.95 \\
p(|x-\mu| < 3\sigma) &\approx 0.997
\end{aligned}$$



Correlation does not imply causation



Correlation does not imply causation



Background: Multivariate Gaussian Distribution

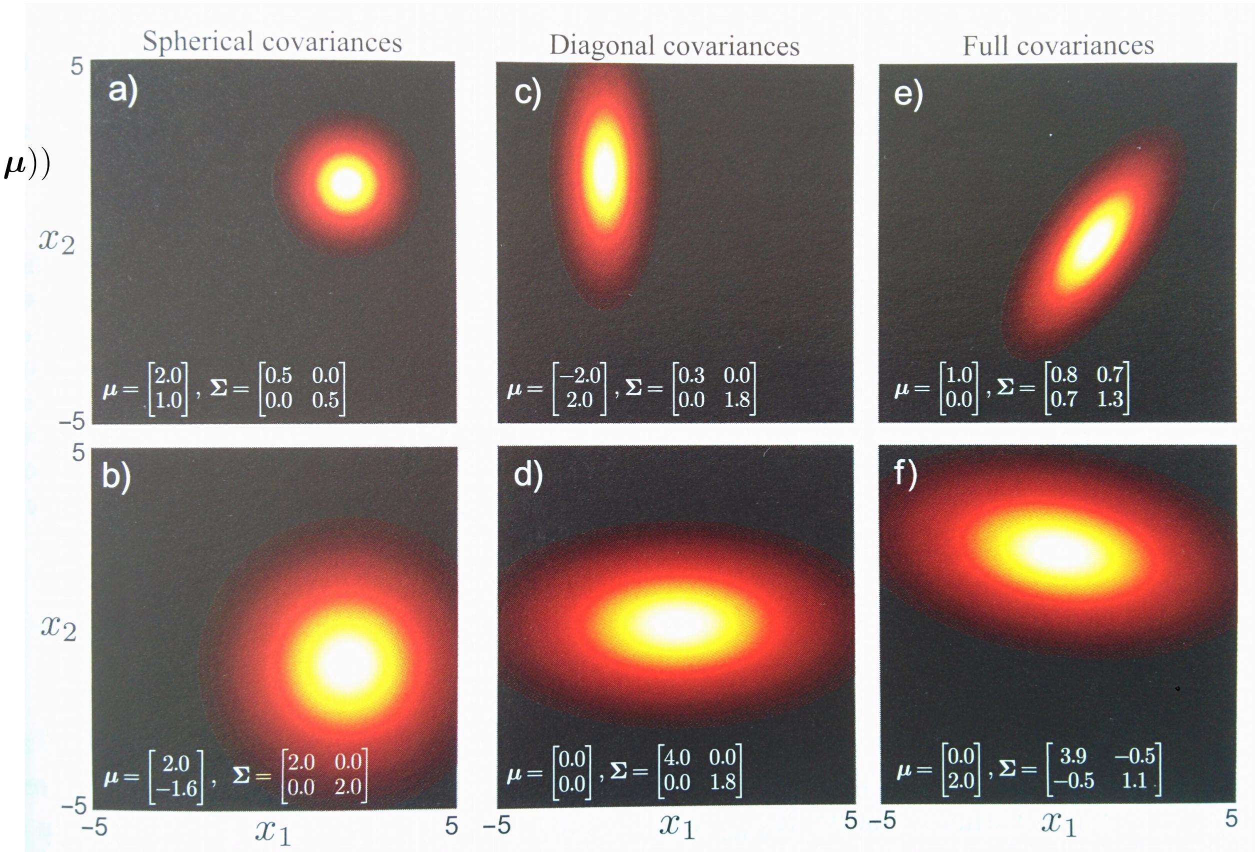
$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\det(\mathbf{\Sigma})^{1/2}}\exp(-0.5(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\det(\mathbf{\Sigma})^{1/2}}\exp(-0.5||\mathbf{x}-\boldsymbol{\mu}||^2_{\mathbf{\Sigma}})$$

For multivariate Gaussians:

$$E[\mathbf{x}] = \boldsymbol{\mu}$$

$$\mathrm{Cov}[\mathbf{x}] = \mathbf{\Sigma}$$



From “Computer Vision: Models, Learning, and Inference” Simon Prince

Background: Multivariate Gaussian Distribution

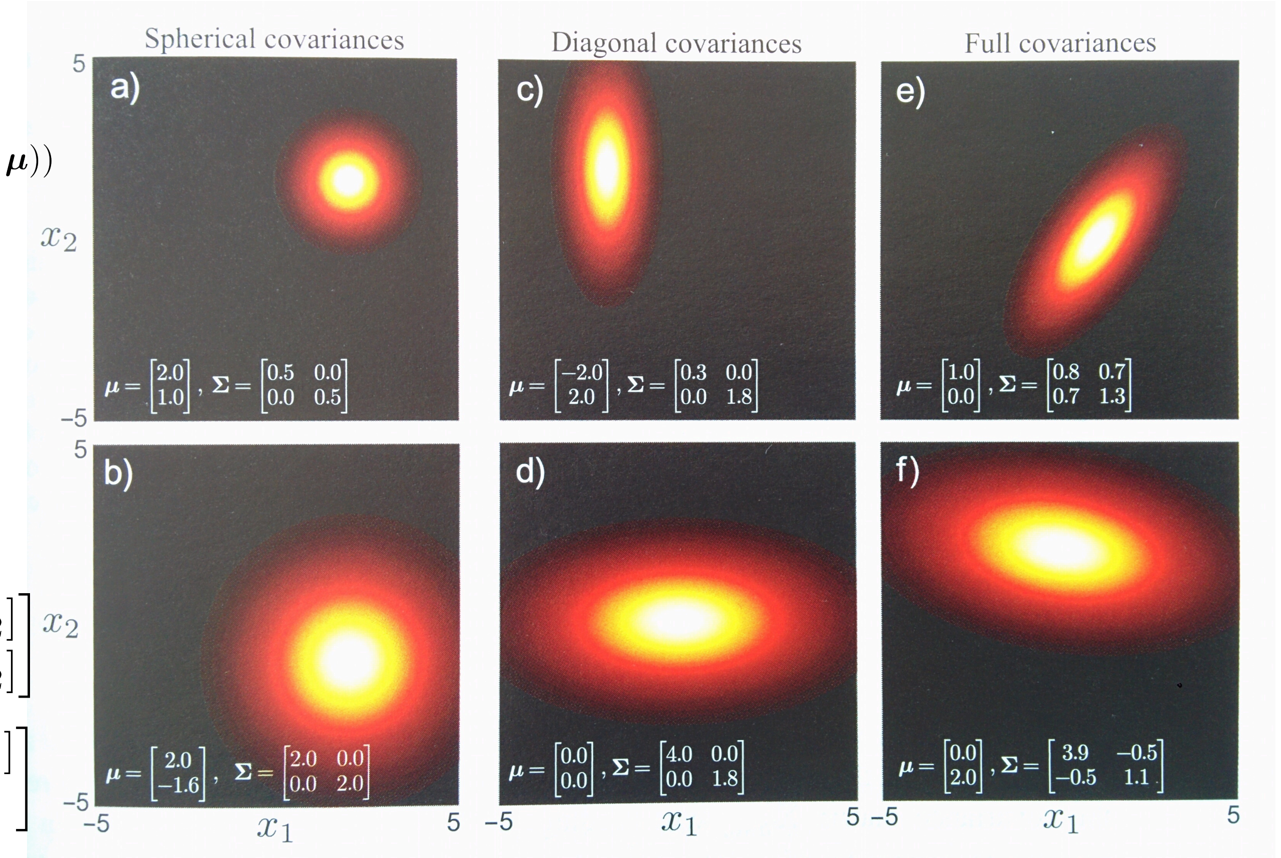
$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\det(\boldsymbol{\Sigma})^{1/2}}\exp(-0.5(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}\det(\boldsymbol{\Sigma})^{1/2}}\exp(-0.5||\mathbf{x}-\boldsymbol{\mu}||^2_{\boldsymbol{\Sigma}})$$

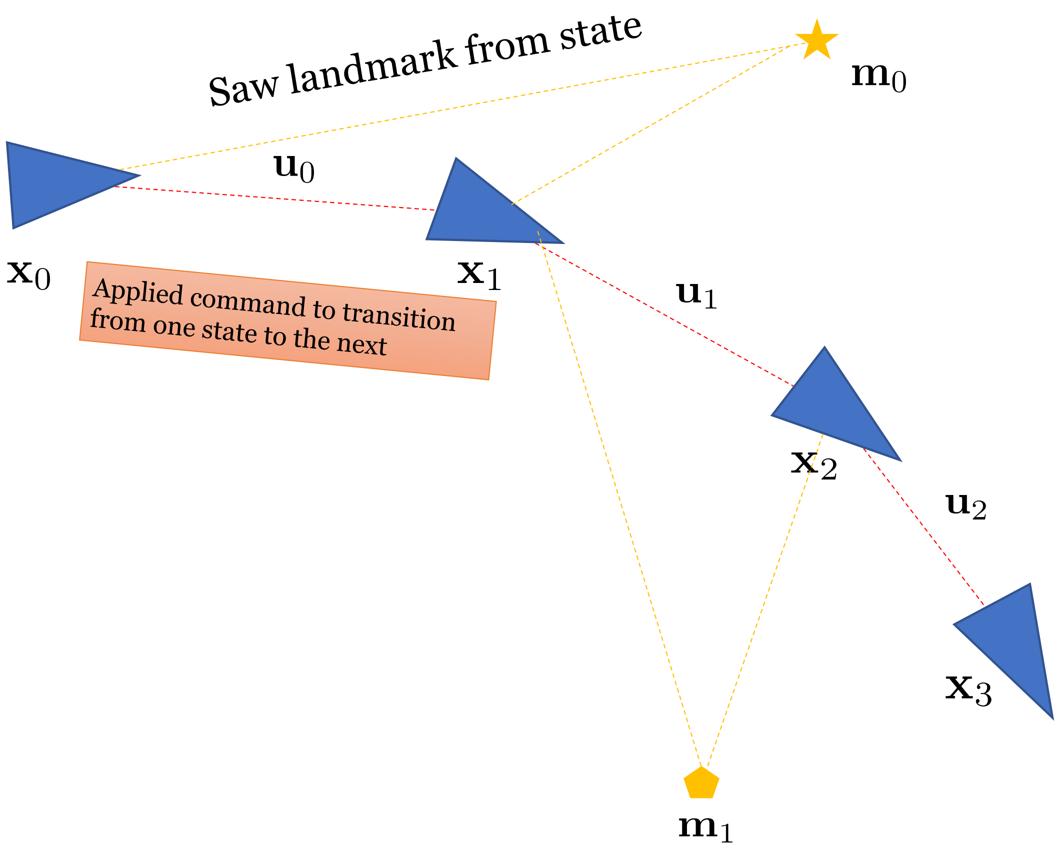
Since we have 2D examples here:

$$\mathrm{Cov}[\mathbf{x}] = \boldsymbol{\Sigma} = \begin{bmatrix}\mathrm{Cov}[x_1,x_1] & \mathrm{Cov}[x_1,x_2]\\ \mathrm{Cov}[x_2,x_1] & \mathrm{Cov}[x_2,x_2]\end{bmatrix}$$

$$= \begin{bmatrix}\mathrm{Var}[x_1] & \mathrm{Cov}[x_1,x_2]\\ \mathrm{Cov}[x_2,x_1] & \mathrm{Var}[x_2]\end{bmatrix}$$



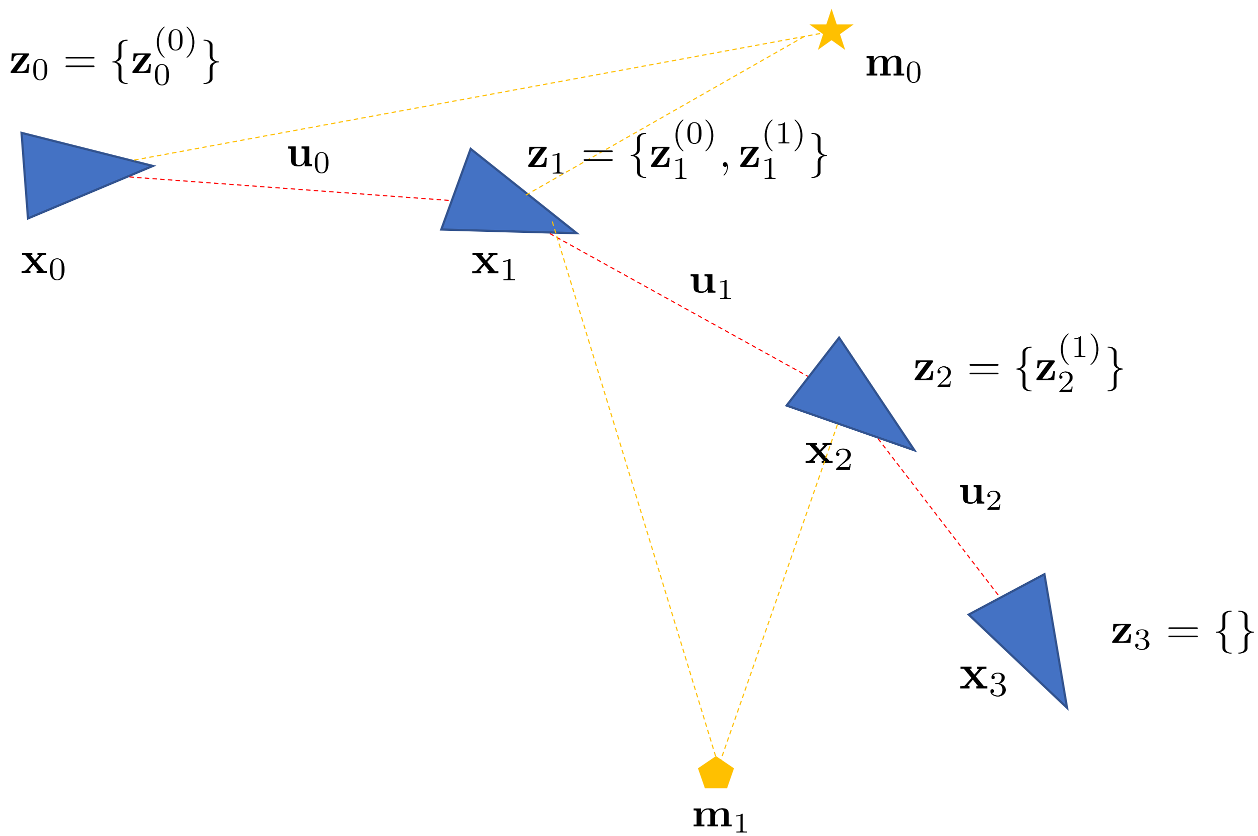
SLAM: graph representation



Map $\mathbf{m} = \{\mathbf{m}_0, \mathbf{m}_1\}$ consists of landmarks that are easily identifiable and cannot be mistaken for one another.

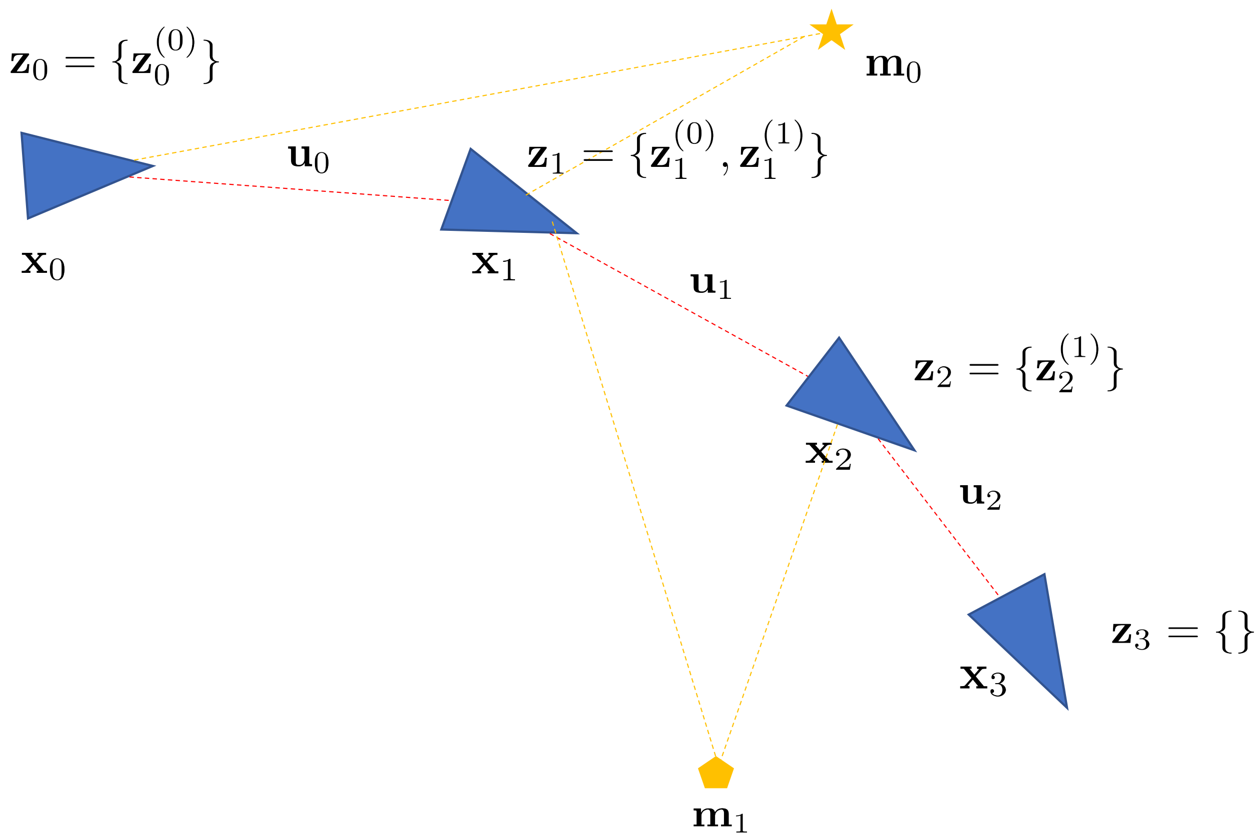
i.e. we are avoiding the data association problem here.

SLAM: graph representation



Map $\mathbf{m} = \{\mathbf{m}_0, \mathbf{m}_1\}$ consists of landmarks that are easily identifiable and cannot be mistaken for one another.

SLAM: graph representation



Notice that the graph is mostly sparse as long as not many states observe the same landmark.

That implies that there are many symbolic dependencies between random variables in

$$p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$$

that are not necessary and can be dropped.

GraphSLAM: SLAM as a Maximum A Posteriori Estimate

Instead of computing the posterior $p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$ we are going to compute its max

See least squares lecture

$$\mathbf{x}^*_{1:T}, \mathbf{m}^* = \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}}\, p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$$

GraphSLAM: SLAM as a Maximum A Posteriori Estimate

Instead of computing the posterior $p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$ we are going to compute its max

$$
\begin{aligned}
\mathbf{x}^*_{1:T}, \mathbf{m}^* &= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \, p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\frac{p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)}{p(\mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)} \right]
\end{aligned}
$$

by definition of conditional distribution

GraphSLAM: SLAM as a Maximum A Posteriori Estimate

Instead of computing the posterior $p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$ we are going to compute its max

$$\begin{aligned}
\mathbf{x}^*_{1:T}, \mathbf{m}^* &= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}}\, p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\frac{p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)}{p(\mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)} \right] \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}}\, p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)
\end{aligned}$$

denominator does not depend on optimization variables

GraphSLAM: SLAM as a Maximum A Posteriori Estimate

Instead of computing the posterior $p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$ we are going to compute its max

$$
\begin{aligned}
\mathbf{x}_{1:T}^*, \mathbf{m}^* &= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}}\, p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\frac{p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)}{p(\mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)} \right] \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}}\, p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0) \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\prod_{t=1}^{T} p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}) \prod_{t=0}^{T} \prod_{\mathbf{z}_t^{(k)} \in \mathbf{z}_t} p(\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k) \right]
\end{aligned}
$$

See Appendix 1 for the derivation of this step

Observation of landmark k at time t

Set of observations that were made at time t

GraphSLAM:
SLAM as a Maximum A Posteriori Estimate

Instead of computing the posterior $p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$ we are going to compute its max

$$
\begin{aligned}
\mathbf{x}^*_{1:T}, \mathbf{m}^* &= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}}\, p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\frac{p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)}{p(\mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)} \right] \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}}\, p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0) \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\prod_{t=1}^{T} \underbrace{p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1})}_{\text{Probabilistic dynamics model}} \prod_{t=0}^{T} \prod_{\mathbf{z}_t^{(k)} \in\, \mathbf{z}_t} \overbrace{p(\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k)}^{\text{Probabilistic sensor measurement model}} \right]
\end{aligned}
$$

GraphSLAM: SLAM as a Maximum A Posteriori Estimate

Instead of computing the posterior $p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$ we are going to compute its max

$$
\begin{aligned}
\mathbf{x}^*_{1:T}, \mathbf{m}^* &= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}}\, p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\frac{p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)}{p(\mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0)} \right] \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}}\, p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1} | \mathbf{x}_0) \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\prod_{t=1}^{T} p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}) \prod_{t=0}^{T} \prod_{\mathbf{z}_t^{(k)} \in\, \mathbf{z}_t} p(\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k) \right] \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\sum_{t=1}^{T} \log p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}) + \sum_{t=0}^{T} \sum_{\mathbf{z}_t^{(k)} \in\, \mathbf{z}_t} \log p(\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k) \right]
\end{aligned}
$$

GraphSLAM: SLAM as a Maximum A Posteriori Estimate

Instead of computing the posterior $p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$ we are going to compute its max

$$\mathbf{x}^*_{1:T}, \mathbf{m}^* = \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\operatorname{argmax}} \left[\sum_{t=1}^{T} \log p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}) + \sum_{t=0}^{T} \sum_{\mathbf{z}_t^{(k)} \in \mathbf{z}_t} \log p(\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k) \right]$$

Main GraphSLAM assumptions:

1. Uncertainty in the dynamics model is Gaussian

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}) + \mathbf{w}_t$$

$$\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$$

so

$$\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1} \sim \mathcal{N}(\mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}), \mathbf{R}_t)$$

2. Uncertainty in the sensor model is Gaussian

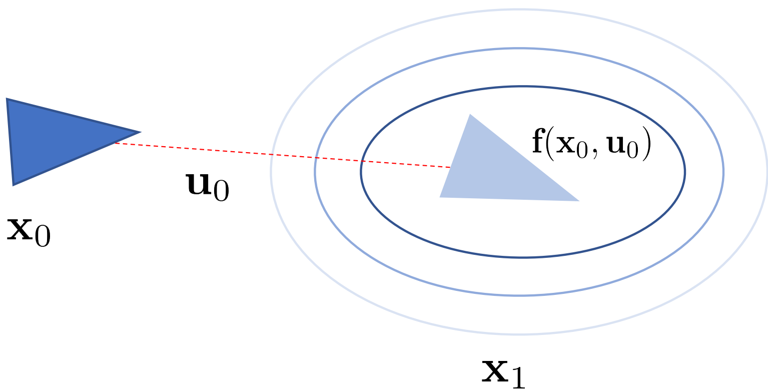
$$\mathbf{z}_t^{(k)} = \mathbf{h}(\mathbf{x}_t, \mathbf{m}_k) + \mathbf{v}_t$$

$$\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$$

so

$$\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k \sim \mathcal{N}(\mathbf{h}(\mathbf{x}_t, \mathbf{m}_k), \mathbf{Q}_t)$$

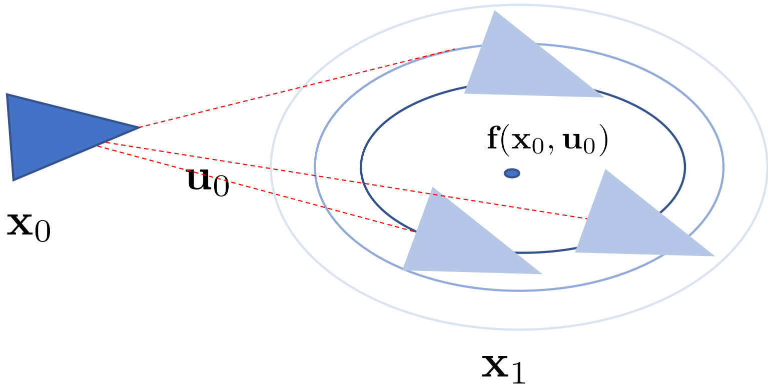
SLAM: noise/errors



$$\mathbf{x}_1 | \mathbf{x}_0, \mathbf{u}_0 \sim \mathcal{N}(\mathbf{f}(\mathbf{x}_0, \mathbf{u}_0), \mathbf{R}_0)$$

Expected to end up at $\mathbf{x}_1 = \mathbf{f}(\mathbf{x}_0, \mathbf{u}_0)$ from $\mathbf{x}_0$

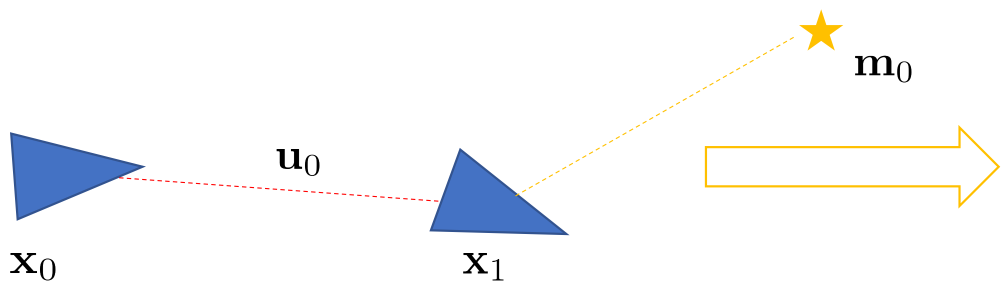
SLAM: noise/errors

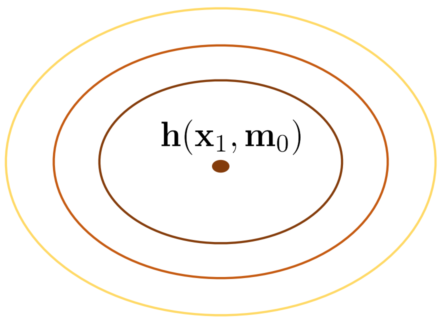


$$\mathbf{x}_1 | \mathbf{x}_0, \mathbf{u}_0 \sim \mathcal{N}(\mathbf{f}(\mathbf{x}_0, \mathbf{u}_0), \mathbf{R}_0)$$

Expected to end up at $\mathbf{x}_1 = \mathbf{f}(\mathbf{x}_0, \mathbf{u}_0)$ from $\mathbf{x}_0$ but we might end up around it, within the ellipse defined by the covariance matrix $\mathbf{R}_0$

SLAM: noise/errors





$$\mathbf{z}_1^{(0)} | \mathbf{x}_1, \mathbf{m}_0 \sim \mathcal{N}(\mathbf{h}(\mathbf{x}_1, \mathbf{m}_0), \mathbf{Q}_1)$$

Expected to get measurement $\mathbf{h}(\mathbf{x}_1, \mathbf{m}_0)$ at state $\mathbf{x}_1$ but it might be somewhere within the ellipse defined by the covariance matrix $\mathbf{Q}_1$

GraphSLAM: SLAM as a least squares problem

Instead of computing the posterior $p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$ we are going to compute its max

$$\mathbf{x}^*_{1:T}, \mathbf{m}^* = \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\text{argmax}} \left[\sum_{t=1}^{T} \log p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}) + \sum_{t=0}^{T} \sum_{\mathbf{z}_t^{(k)} \in \mathbf{z}_t} \log p(\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k) \right]$$

$$= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\text{argmax}} \left[-\sum_{t=1}^{T} ||\mathbf{x}_t - \mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1})||^2_{\mathbf{R}_t} - \sum_{t=0}^{T} \sum_{\mathbf{z}_t^{(k)} \in \mathbf{z}_t} ||\mathbf{z}_t^{(k)} - \mathbf{h}(\mathbf{x}_t, \mathbf{m}_k)||^2_{\mathbf{Q}_t} \right]$$

Notation:

$\mathbf{x}^T \mathbf{Q}^{-1} \mathbf{x} = ||\mathbf{x}||^2_{\mathbf{Q}}$

$\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1} \sim \mathcal{N}(\mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}), \mathbf{R}_t)$

$\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k \sim \mathcal{N}(\mathbf{h}(\mathbf{x}_t, \mathbf{m}_k), \mathbf{Q}_t)$

GraphSLAM: SLAM as a least squares problem

Instead of computing the posterior $p(\mathbf{x}_{1:T}, \mathbf{m} \mid \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)$ we are going to compute its max

$$\begin{aligned}
\mathbf{x}^*_{1:T}, \mathbf{m}^* &= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\text{argmax}} \left[\sum_{t=1}^{T} \log p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}) + \sum_{t=0}^{T} \sum_{\mathbf{z}_t^{(k)} \in \mathbf{z}_t} \log p(\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k) \right] \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\text{argmax}} \left[-\sum_{t=1}^{T} ||\mathbf{x}_t - \mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1})||^2_{\mathbf{R}_t} - \sum_{t=0}^{T} \sum_{\mathbf{z}_t^{(k)} \in \mathbf{z}_t} ||\mathbf{z}_t^{(k)} - \mathbf{h}(\mathbf{x}_t, \mathbf{m}_k)||^2_{\mathbf{Q}_t} \right] \\
&= \underset{\mathbf{x}_{1:T}, \mathbf{m}}{\text{argmin}} \left[\sum_{t=1}^{T} ||\mathbf{x}_t - \mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1})||^2_{\mathbf{R}_t} + \sum_{t=0}^{T} \sum_{\mathbf{z}_t^{(k)} \in \mathbf{z}_t} ||\mathbf{z}_t^{(k)} - \mathbf{h}(\mathbf{x}_t, \mathbf{m}_k)||^2_{\mathbf{Q}_t} \right]
\end{aligned}$$

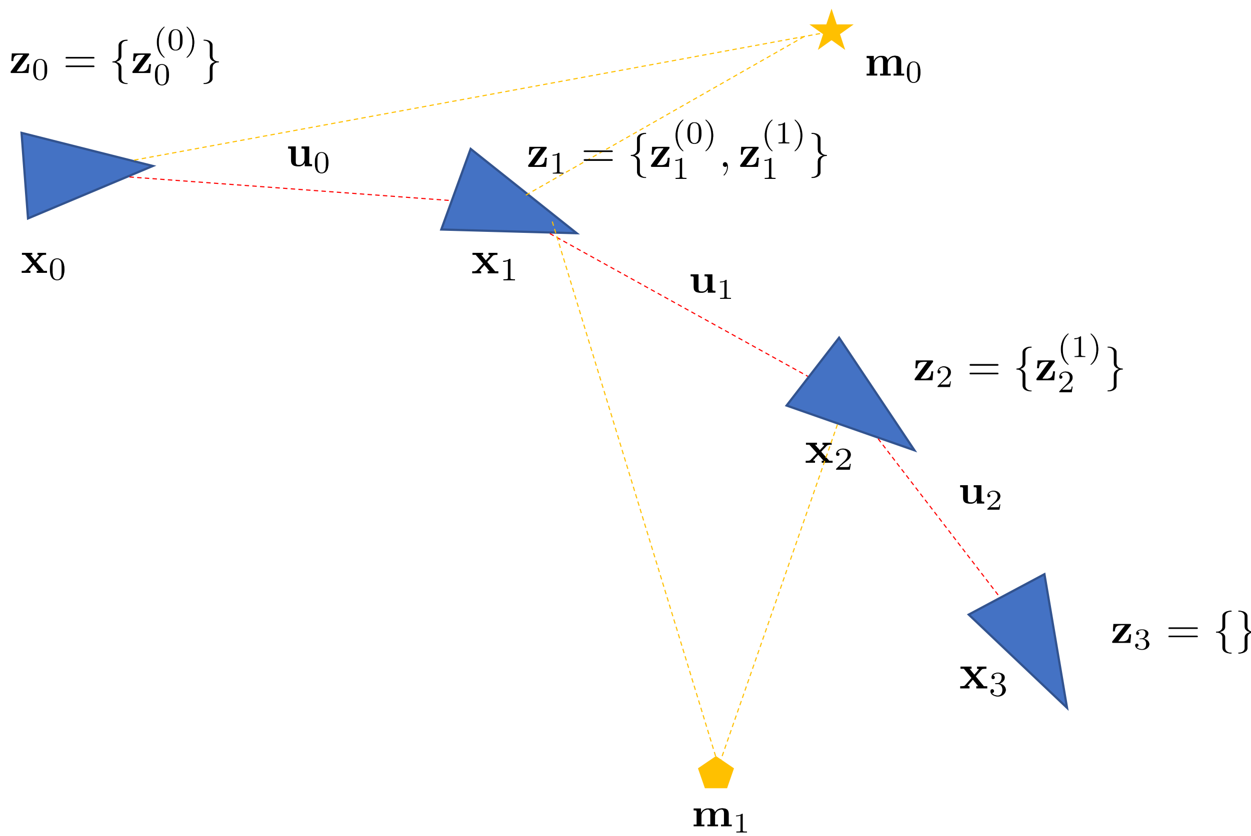
Notation:

$\mathbf{x}^T \mathbf{Q}^{-1} \mathbf{x} = ||\mathbf{x}||^2_{\mathbf{Q}}$

$\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1} \sim \mathcal{N}(\mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}), \mathbf{R}_t)$

$\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k \sim \mathcal{N}(\mathbf{h}(\mathbf{x}_t, \mathbf{m}_k), \mathbf{Q}_t)$

GraphSLAM: example



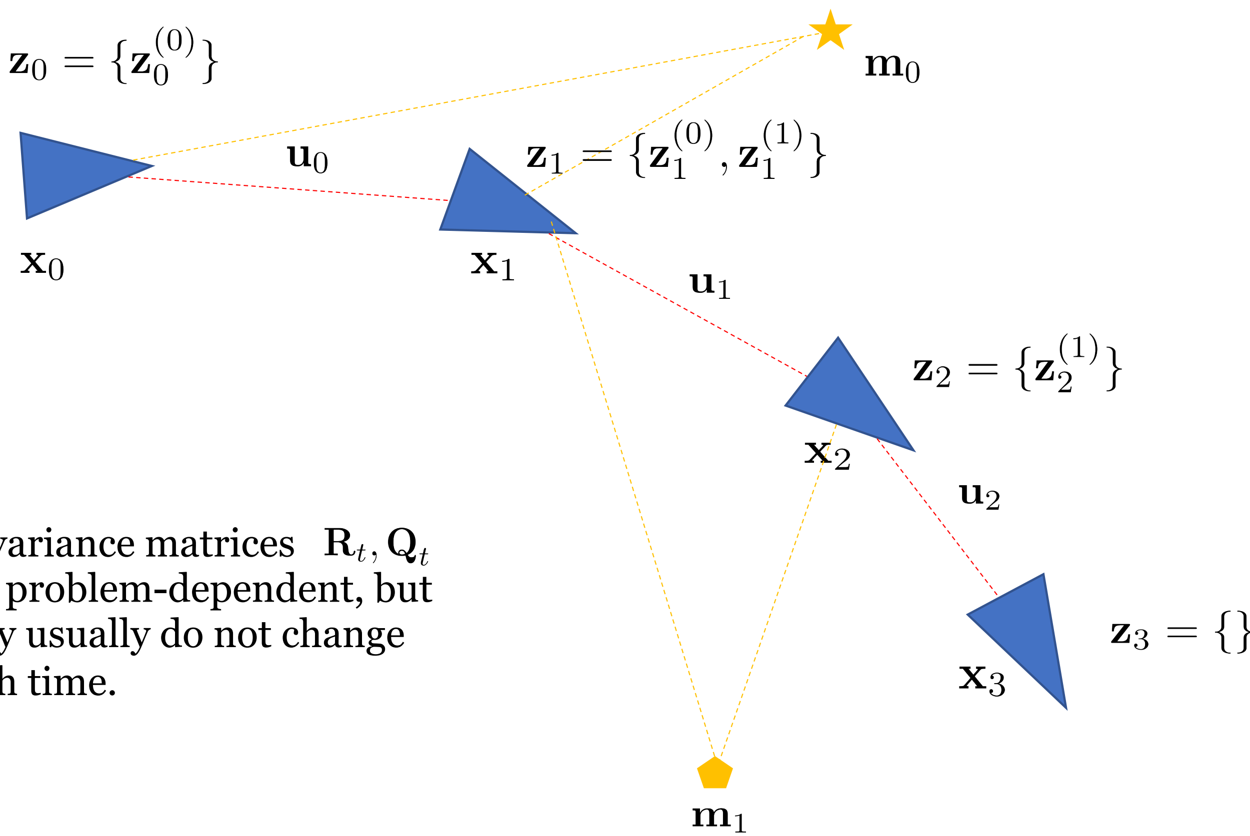
Need to minimize the sum of the following quadratic terms:

$$||\mathbf{x}_1 - \mathbf{f}(\mathbf{x}_0, \mathbf{u}_0)||^2_{\mathbf{R}_1}$$
$$||\mathbf{x}_2 - \mathbf{f}(\mathbf{x}_1, \mathbf{u}_1)||^2_{\mathbf{R}_2}$$
$$||\mathbf{x}_3 - \mathbf{f}(\mathbf{x}_2, \mathbf{u}_2)||^2_{\mathbf{R}_3}$$
$$||\mathbf{z}_0^{(0)} - \mathbf{h}(\mathbf{x}_0, \mathbf{m}_0)||^2_{\mathbf{Q}_0}$$
$$||\mathbf{z}_1^{(0)} - \mathbf{h}(\mathbf{x}_1, \mathbf{m}_0)||^2_{\mathbf{Q}_1}$$
$$||\mathbf{z}_1^{(1)} - \mathbf{h}(\mathbf{x}_1, \mathbf{m}_1)||^2_{\mathbf{Q}_1}$$
$$||\mathbf{z}_2^{(1)} - \mathbf{h}(\mathbf{x}_2, \mathbf{m}_1)||^2_{\mathbf{Q}_2}$$

with respect to variables:
$\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{m}_0 \quad \mathbf{m}_1$

initial state $\mathbf{x}_0$ is given

GraphSLAM: example



Covariance matrices $\mathbf{R}_t, \mathbf{Q}_t$ are problem-dependent, but they usually do not change with time.

Need to minimize the sum of the following quadratic terms:

$$||\mathbf{x}_1 - \mathbf{f}(\mathbf{x}_0, \mathbf{u}_0)||^2_{\mathbf{R}_1}$$
$$||\mathbf{x}_2 - \mathbf{f}(\mathbf{x}_1, \mathbf{u}_1)||^2_{\mathbf{R}_2}$$
$$||\mathbf{x}_3 - \mathbf{f}(\mathbf{x}_2, \mathbf{u}_2)||^2_{\mathbf{R}_3}$$
$$||\mathbf{z}_0^{(0)} - \mathbf{h}(\mathbf{x}_0, \mathbf{m}_0)||^2_{\mathbf{Q}_0}$$
$$||\mathbf{z}_1^{(0)} - \mathbf{h}(\mathbf{x}_1, \mathbf{m}_0)||^2_{\mathbf{Q}_1}$$
$$||\mathbf{z}_1^{(1)} - \mathbf{h}(\mathbf{x}_1, \mathbf{m}_1)||^2_{\mathbf{Q}_1}$$
$$||\mathbf{z}_2^{(1)} - \mathbf{h}(\mathbf{x}_2, \mathbf{m}_1)||^2_{\mathbf{Q}_2}$$

with respect to variables:

$\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{m}_0 \quad \mathbf{m}_1$

initial state $\mathbf{x}_0$ is given

Examples of dynamics and sensor models

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}) + \mathbf{w}_t$$

Can be any of the dynamical systems we saw in Lecture 2.

$$\mathbf{z}_t^{(k)} = \mathbf{h}(\mathbf{x}_t, \mathbf{m}_k) + \mathbf{v}_t$$

$\mathbf{z}_t^{(k)}$ can be any of the sensors we saw in Lecture 4:

- Laser scan $\{(r_i, \theta_i)\}_{i=1:K}$ where $\mathbf{m}_k$ is an occupancy grid
- Range and bearing (r, θ) to the landmark $\mathbf{m}_k = (x_k, y_k, z_k)$
- Bearing measurements from images
- Altitude/Depth
- Gyroscope
- Accelerometer

Appendix 1

Claim:

$$p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \quad = \quad p(\mathbf{x}_0) \prod_{t=1}^{T} p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1}) \prod_{t=0}^{T} \prod_{\mathbf{z}_t^{(k)} \in \, \mathbf{z}_t} p(\mathbf{z}_t^{(k)} | \mathbf{x}_t, \mathbf{m}_k)$$

Proof:

$$\begin{aligned}
p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) &= p(\mathbf{z}_T | \mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T-1}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)\; p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T-1}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= p(\mathbf{z}_T | \mathbf{x}_T, \mathbf{m})\; p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T-1}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= p(\mathbf{z}_T | \mathbf{x}_T, \mathbf{m})\; p(\mathbf{z}_{T-1} | \mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T-2}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)\; p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T-2}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= p(\mathbf{z}_T | \mathbf{x}_T, \mathbf{m})\; p(\mathbf{z}_{T-1} | \mathbf{x}_{T-1}, \mathbf{m})\; p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{z}_{0:T-2}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&\ldots \\
&= \prod_{t=0}^{T} p(\mathbf{z}_t | \mathbf{x}_t, \mathbf{m})\; p(\mathbf{x}_{1:T}, \mathbf{m}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= \prod_{t=0}^{T} p(\mathbf{z}_t | \mathbf{x}_t, \mathbf{m})\; p(\mathbf{x}_T | \mathbf{x}_{1:T-1}, \mathbf{m}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)\;\; p(\mathbf{x}_{1:T-1}, \mathbf{m}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= \prod_{t=0}^{T} p(\mathbf{z}_t | \mathbf{x}_t, \mathbf{m})\; p(\mathbf{x}_T | \mathbf{x}_{T-1}, \mathbf{u}_{T-1})\; p(\mathbf{x}_{1:T-1}, \mathbf{m}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&= \prod_{t=0}^{T} p(\mathbf{z}_t | \mathbf{x}_t, \mathbf{m})\; p(\mathbf{x}_T | \mathbf{x}_{T-1}, \mathbf{u}_{T-1})\; p(\mathbf{x}_{T-1} | \mathbf{x}_{T-2}, \mathbf{m}, \mathbf{u}_{0:T-1}, \mathbf{x}_0)\; p(\mathbf{x}_{1:T-2}, \mathbf{m}, \mathbf{u}_{0:T-1}, \mathbf{x}_0) \\
&\ldots \\
&= \left[\prod_{t=0}^{T} p(\mathbf{z}_t | \mathbf{x}_t, \mathbf{m})\right]\; p(\mathbf{x}_0) \prod_{t=1}^{T} p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_{t-1})
\end{aligned}$$

Appendix 1

Claim: $$p(\mathbf{z}_t|\mathbf{x}_t,\mathbf{m}) = \prod_{\mathbf{z}_t^{(k)}\in\mathbf{z}_t} p(\mathbf{z}_t^{(k)}|\mathbf{x}_t,\mathbf{m}_k)$$ where $$\mathbf{z}_t = \{\mathbf{z}_t^{(k)} \text{ iff landmark } \mathbf{m}_k \text{ was observed}\}$$ $$\mathbf{m} = \{\text{landmarks } \mathbf{m}_k\}$$

Proof:

Suppose without loss of generality that $\mathbf{z}_t = \{\mathbf{z}_t^{(k)},\ k = 1...K\}$ and $\mathbf{m} = \{\mathbf{m}_k,\ k = 1...K\}$
i.e. that all landmarks were observed from the state at time t. Then:

$$\begin{aligned}
p(\mathbf{z}_t^{(1)},...,\mathbf{z}_t^{(K)}|\mathbf{x}_t,\mathbf{m}) &= p(\mathbf{z}_t^{(1)}|\mathbf{z}_t^{(2)},...,\mathbf{z}_t^{(K)},\mathbf{x}_t,\mathbf{m})\ p(\mathbf{z}_t^{(2)},...,\mathbf{z}_t^{(K)}|\mathbf{x}_t,\mathbf{m}) \\
&= p(\mathbf{z}_t^{(1)}|\mathbf{x}_t,\mathbf{m}_1)\ p(\mathbf{z}_t^{(2)},...,\mathbf{z}_t^{(K)}|\mathbf{x}_t,\mathbf{m}) \\
&= p(\mathbf{z}_t^{(1)}|\mathbf{x}_t,\mathbf{m}_1)\ p(\mathbf{z}_t^{(2)}|\mathbf{z}_t^{(3)},...,\mathbf{z}_t^{(K)},\mathbf{x}_t,\mathbf{m})\ p(\mathbf{z}_t^{(3)},...,\mathbf{z}_t^{(K)}|\mathbf{x}_t,\mathbf{m}) \\
&= p(\mathbf{z}_t^{(1)}|\mathbf{x}_t,\mathbf{m}_1)\ p(\mathbf{z}_t^{(2)}|\mathbf{x}_t,\mathbf{m}_2)\ p(\mathbf{z}_t^{(3)},...,\mathbf{z}_t^{(K)}|\mathbf{x}_t,\mathbf{m}) \\
&\dots \\
&= \prod_{k=1}^{K} p(\mathbf{z}_t^{(k)}|\mathbf{x}_t,\mathbf{m}_k)
\end{aligned}$$