Probability and Statistics Bases for Robotics

Javier González Jiménez

Reference Books:

- **State Estimation For Robotics**. Timothy D. Barfoot. Cambridge University Press. 2019. (Available online)
- **Computer vision: models, learning and inference**. Simon Prince. Cambridge University Press 2012
- **Introduction to Probability.** Charles Grinstead and Laurie Snell**. A GNU book.**
- **Simultaneous Localization and Mapping for Mobile Robots: Introduction and Methods.** Juan-Antonio Fernández-Madrigal and José Luis Blanco Claraco. IGI-Global. 2013. (In our school Library)

Content

- Probability principles:

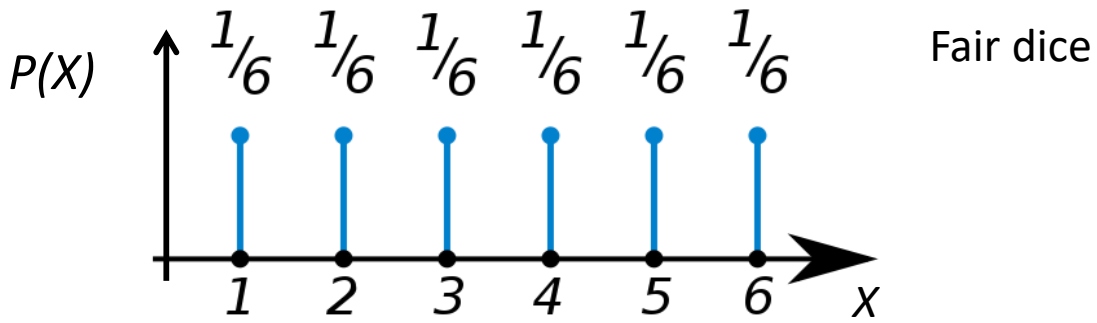
  Probability functions, joint, conditional, marginalization, independence, graphical model, Markov assumption, expectation
- Bayes rule. Recursive Bayes
- Gaussian distribution
- Uncertainty propagation
- Sampling from a distribution
- Parameter estimation
- Approaching probabilistic robotics

Discrete Random Variables

- a random variable (RV) is a variable whose values depend on outcomes of a random phenomenon
- The discrete RV (or event) *X* can take on a countable number of values in $\{x_1, x_2, \ldots, x_n\}$
- $P(X=x_i)$ (or $P(x_i)$, for short) is the probability that the random variable *X* takes on value x_i
- *P(X)* is called probability [mass] function

Example: Event *X* ="Rolling a dice"

$x_i \in \{1, 2, 3, 4, 5, 6\}$



Fair dice

Discrete Random Variables

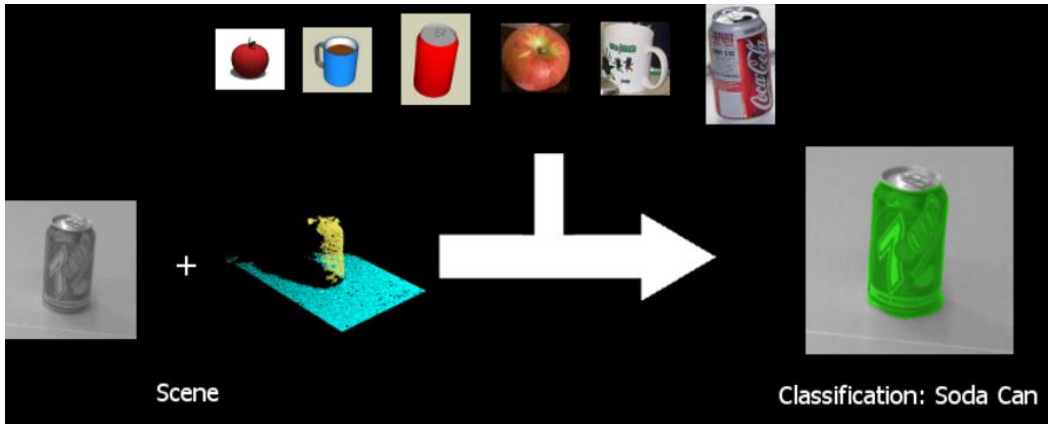
Some robotic examples

- Localization in a grid

Event “what cell is the robot in”

$P(X=cell(i,j))$

- Recognizing an object



Event “what object is it”

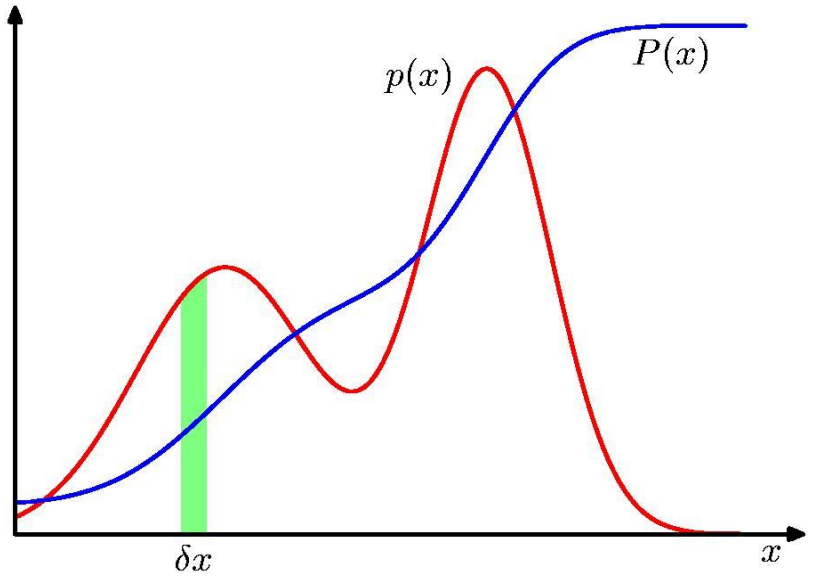
$P(X=x_i)$ $x_i \in$ {apple, mug, can, tomato, …}

(Courtesy: cs.washington.edu)

Continuous Random Variables

- *X* takes on values in the continuum
- *p(X=x)* is a probability **density** function (pdf):

is NOT the probability that *X* takes the value *x*



probability

pdf

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

$$P(x \leq z) = \int_{-\infty}^{z} p(x)dx$$

Cumulative distribution function (a probabiity)

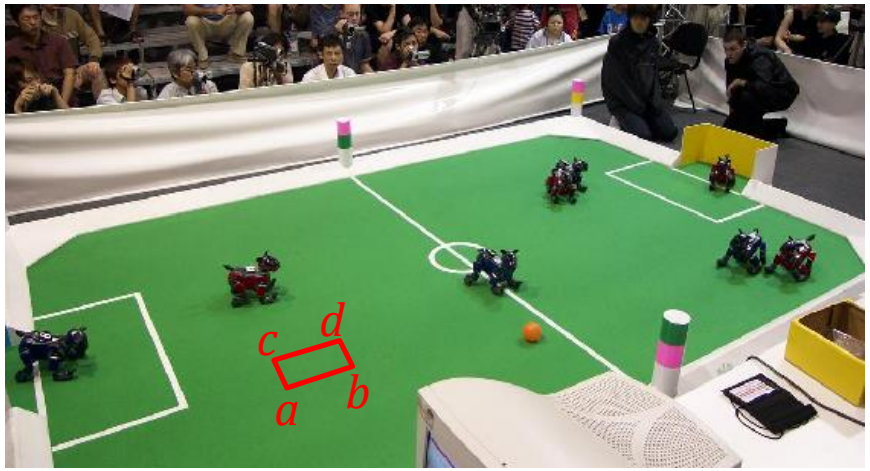
Notation clarification: We will use P for discrete random variables and p for continuous

When an expression applies to both, either one can be used. Sometimes we are using P_r instead of P

Continuous Random Variables

Some robotic examples

- Localization in a continuous plane



- Measuring the distance to an object

- The probability that the robot is at a given position is cero
- The probability that the robot position (x,y) is in the rectangle [a,b] [c,d]

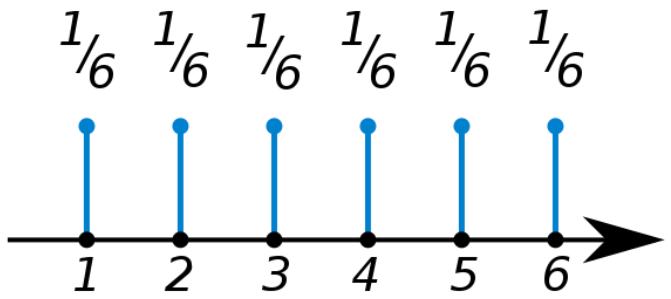
$$P(x \in [a,b], y \in [c,d]) = \int_a^b \int_c^d p(x,y)dxdy\,dxdy$$

The probability that the laser distance r is in the interval [a,b] is:

$$P(r \in [a,b]) = \int_a^b p(r)dr$$

Then, we can find two kinds of distributions

Mass probability functions: Random variable is discrete

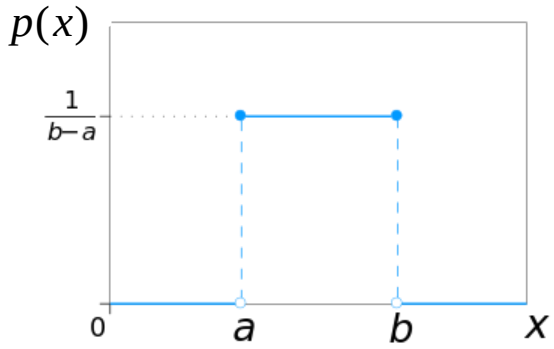


$$P(x) = P(X = x \in S)$$

$$\sum_{x \in S} P(x) = 1$$

Example: Rolling a fair dice $S = \{1, 2, 3, 4, 5, 6\}$

Probability density function (pdf) : Random variable is continuous



Uniform pdf in the interval [a,b]

$$p(x) \geq 0$$

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

Axioms of Probability Theory

P(A) denotes probability that proposition *A* is true (e.g. $X=x_i$)

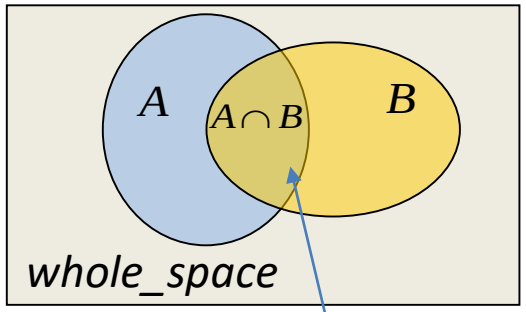
$$0 \le P(A) \le 1$$

$$P(whole\_space) = 1$$

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

Joint probability

A, B are outcomes with the same whole space



$A \cap B = 0$ if A and B mutually exclusive (not possible that both occur at the same time, e.g. head/tail when flipping a coin)

Example: Rolling a dice

A: outcome "odd number" $P(A) = 3/6 = 1/2$

B: outcome {4,5,6} $P(B) = 3/6 = 1/2$

$A \cap B$: outcome {5} $P(A \cap B) = 1/6$

$$P(A \cup B) = P(\{1,3,4,5,6\}) = 5/6$$

$$P(A \cup B) = 1/2 + 1/2 - 1/6 = 5/6$$

Joint probability

Probability of two events X, Y:

$P(x \cap y) = P(x, y)$ means $P(X = x \; and \; Y = y)$

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

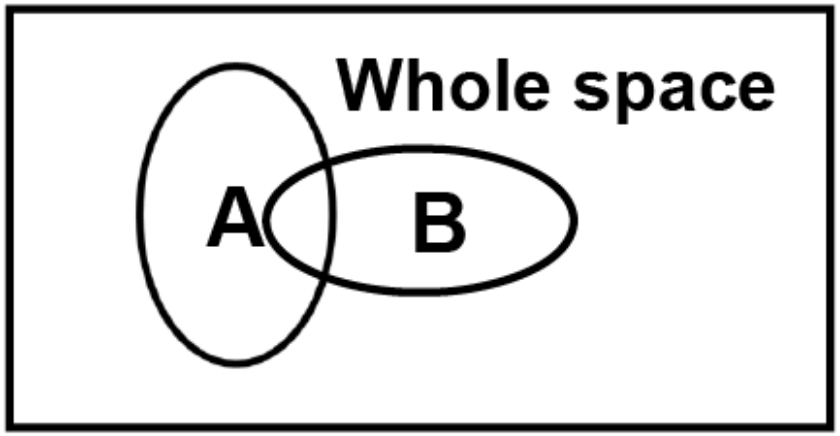
Known as the Product rule

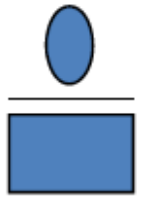
Conditional probability

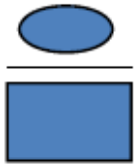
probability the event X takes the value x

given that Y has ALREADY taken the value y

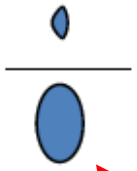
A way to "visualize" probabilities (from a frequentist perspective)



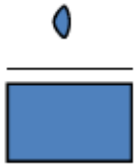
$P(A) =$ 

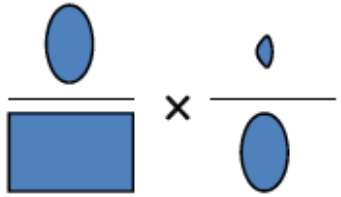
$P(B) =$ 

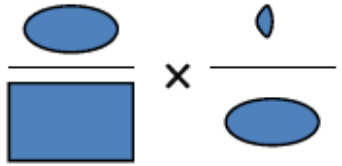
$P(A|B) =$ 

$P(B|A) =$ 

The whole space now is A

$P(A \cap B) =$ 

$P(A) \times P(B|A) =$  $= P(A \cap B)$

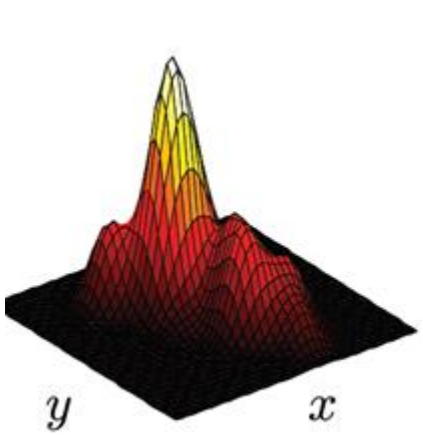
$P(B) \times P(A|B) =$  $= P(A \cap B)$

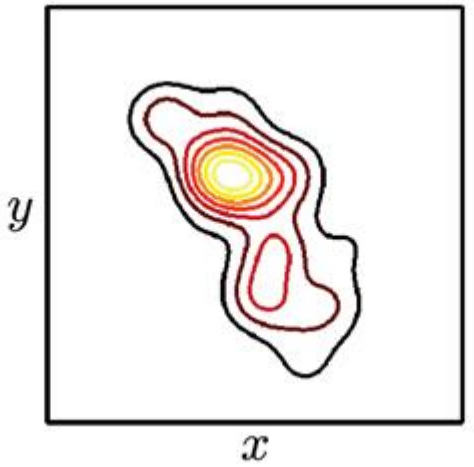
Courtesy: Bob Cousins, HCPSS, 2009

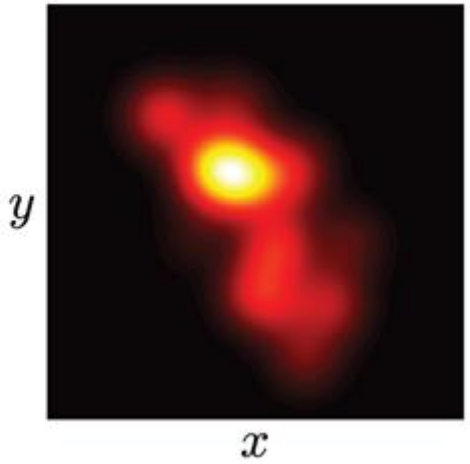
Joint Probability

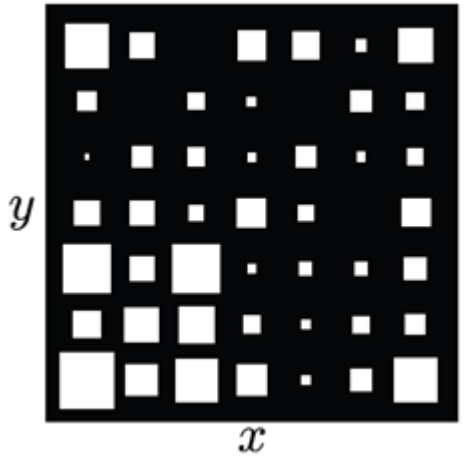
A more general visualization: the joint probability space

x, y continuous
(3 ways of visualization)

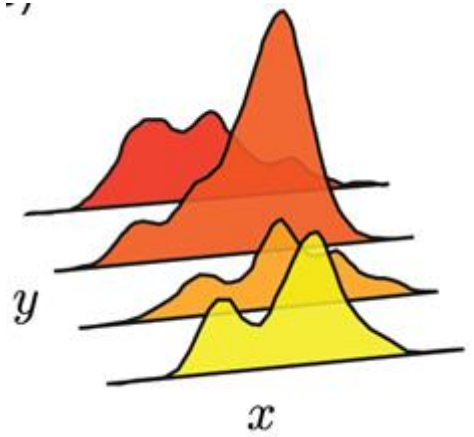




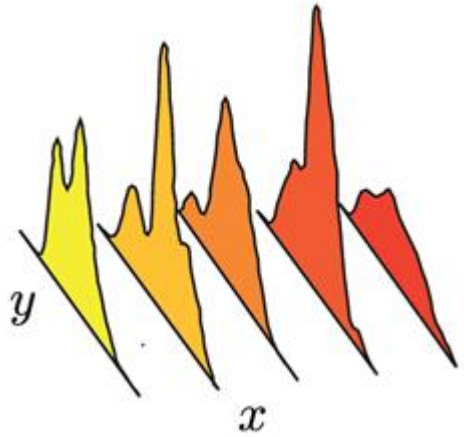




x, y discrete



x continuous
y discrete



x discrete
y continuous

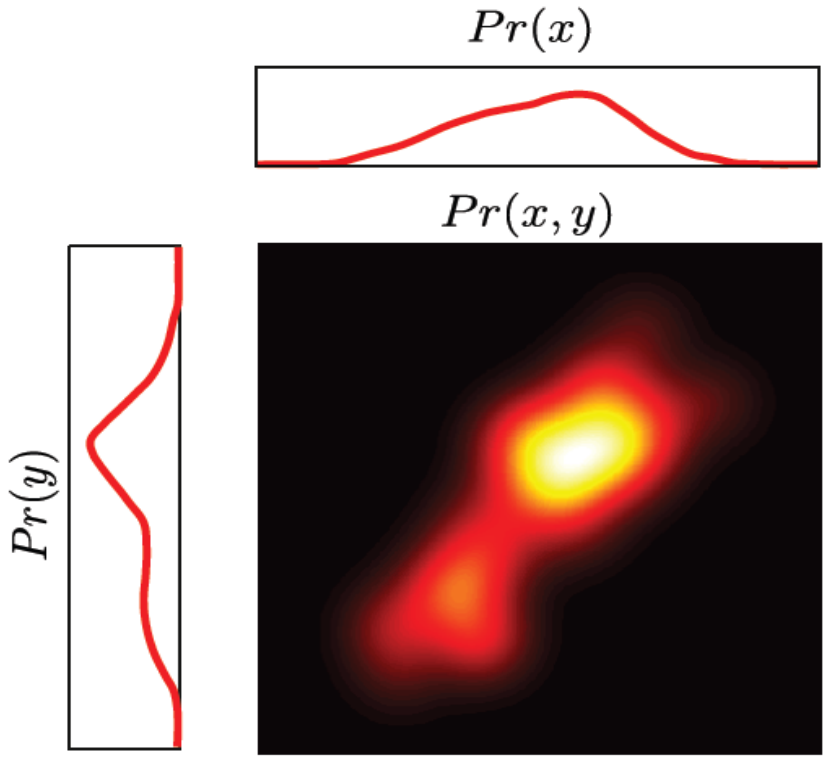
(Courtesy: Prince, 2012)

Marginalization

Recovers the probability distribution of any variable in a joint distribution by integrating (or summing) over the other variables

$$Pr(x) = \int_y P_r(x, y)dy$$ (y continuous)

$$Pr(x) = \sum_y P_r(x, y)$$ (y discrete)



Marginalizing out y

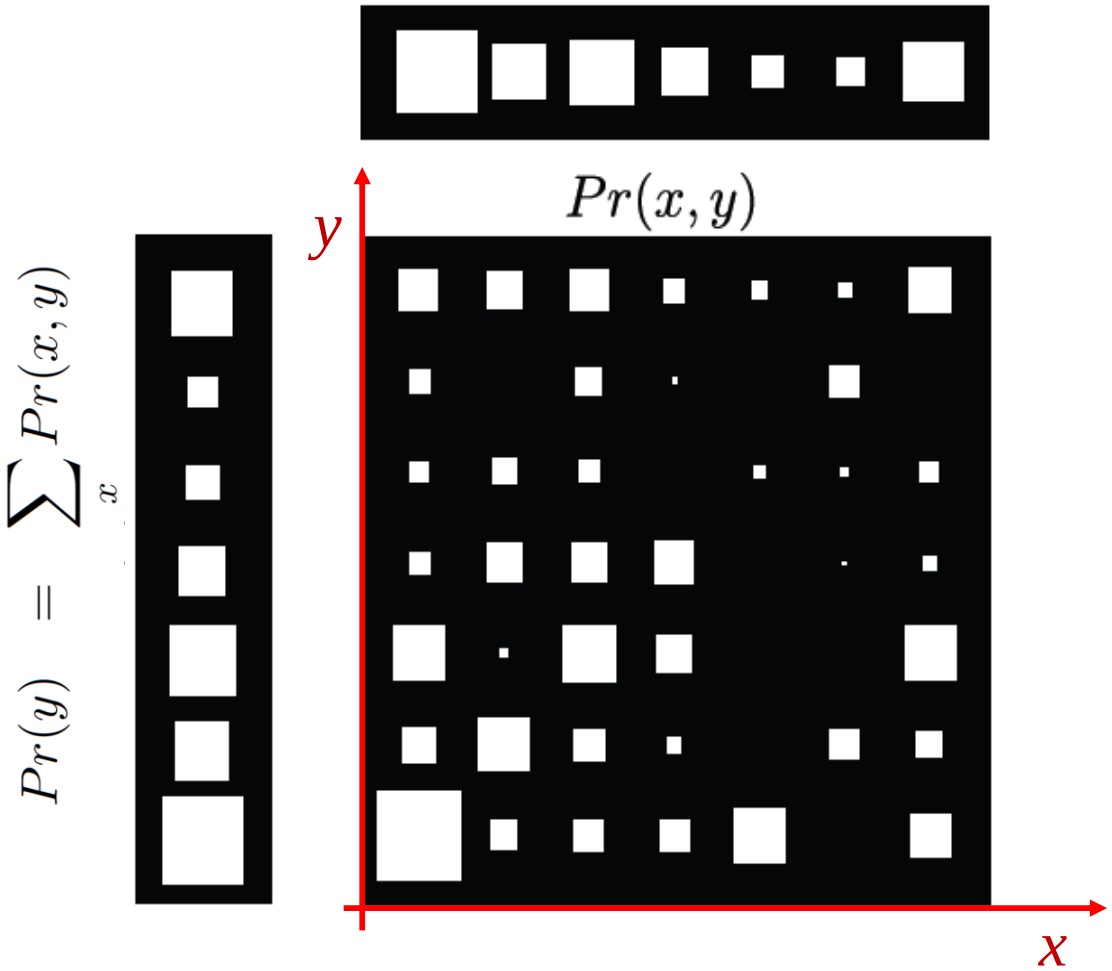
Marginalizing out x

(Courtesy: Prince, 2012)

Marginalization

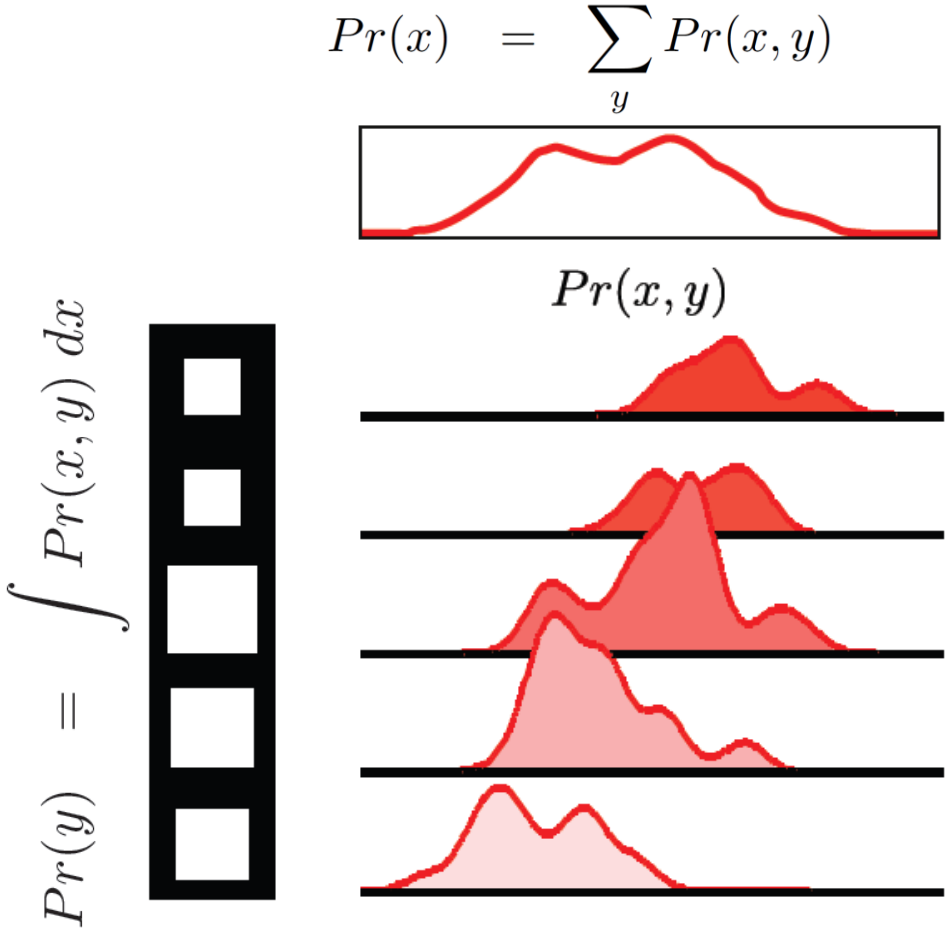
x, y discrete

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



x continuous, y discrete

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

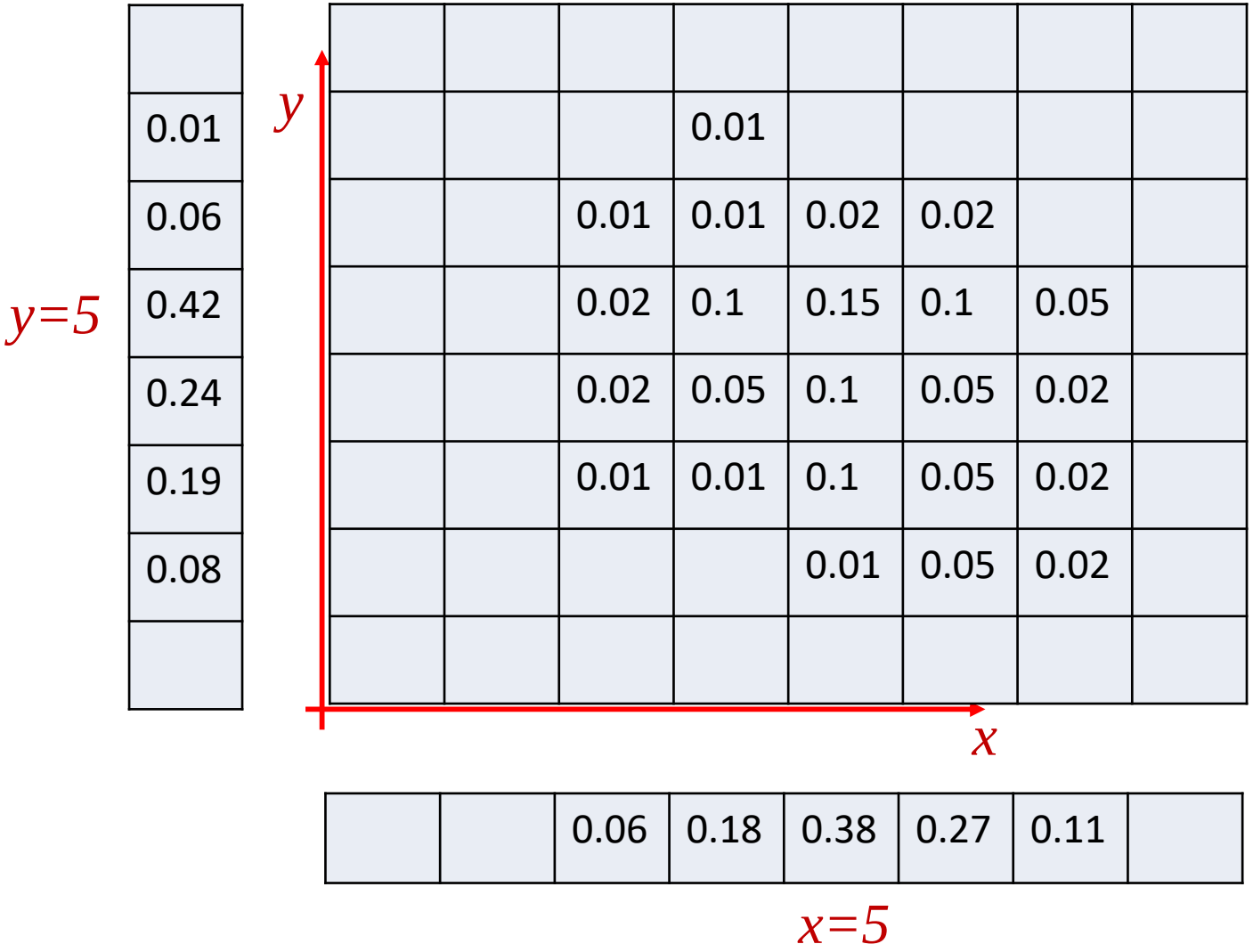


(Courtesy: Prince, 2012)

Example: Distribution of the position of a robot in a grid

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

x, y discrete



Axiom of a probability distribution:

$$\sum_x \sum_y P(x, y) = 1$$

Probability that the robot is in the cell (5,5)

$$P(5,5) = 0.15$$

Probability that the robot is in the column 5

$$P(x = 5) = 0.38$$

Probability that the robot is between the column 4 and 6

$$P(4 \leq x \leq 6) = 0.62$$

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

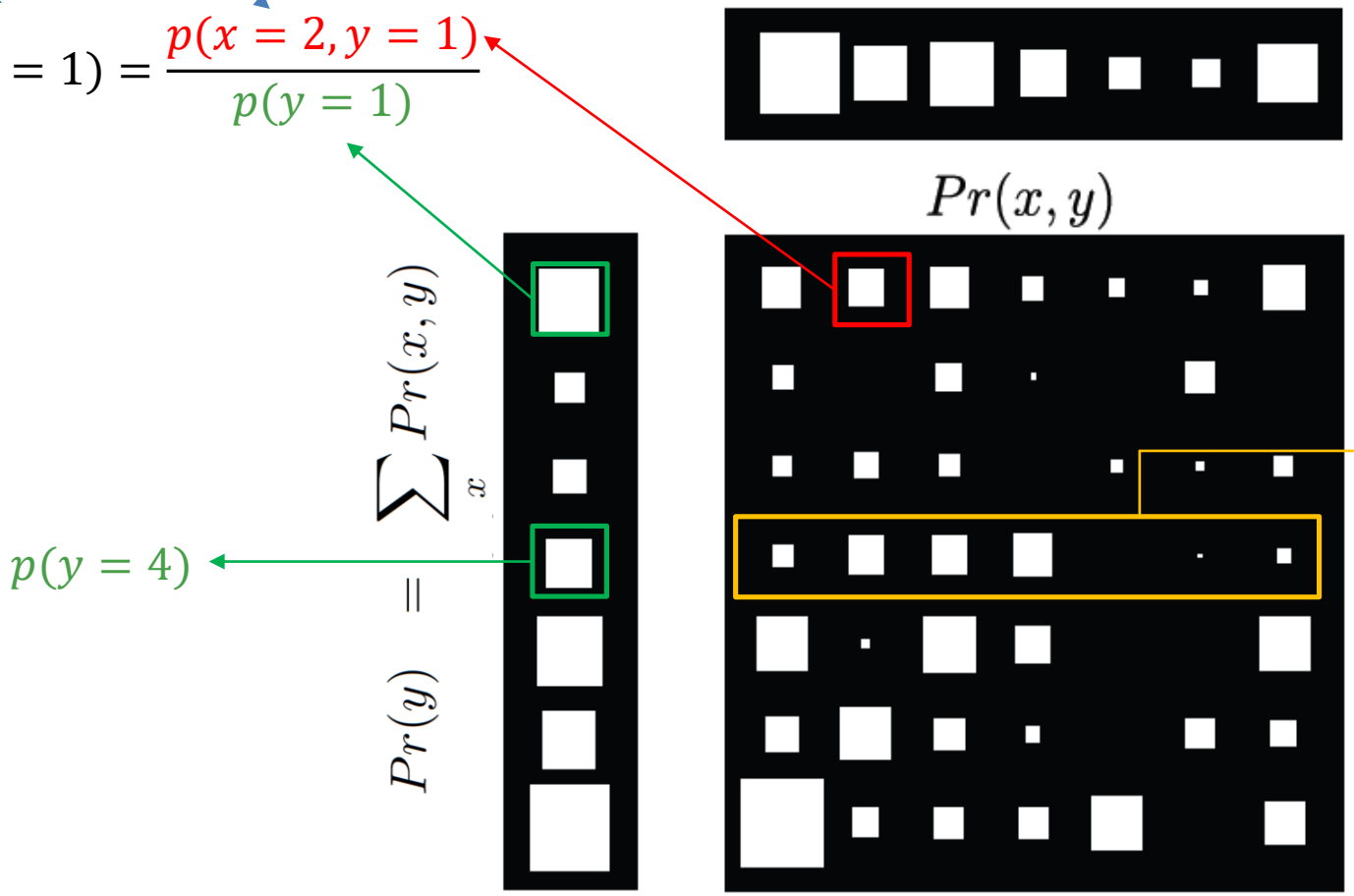
Conditional

Same distribution but scaled by $p(y = 1)$

x, y discrete

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

$$p(x = 2|y = 1) = \frac{p(x = 2, y = 1)}{p(y = 1)}$$



$$p(x|y = 4) = \frac{p(x, y = 4)}{p(y = 4)}$$

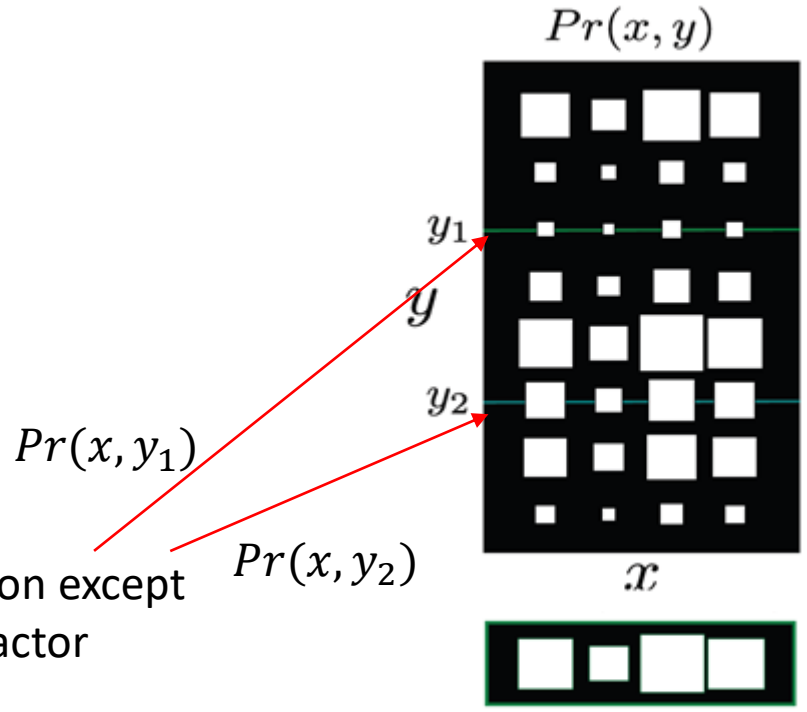
(Courtesy: Prince, 2012)

Independence

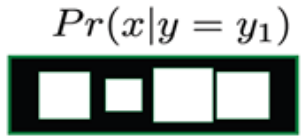
If two variables x and y are independent then the variable x tells us nothing about the variable y (and vice-versa)

$$Pr(x|y) = Pr(x)$$
$$Pr(y|x) = Pr(y)$$

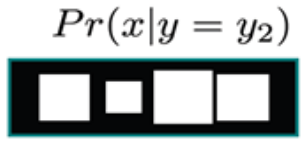
Example for x, y *independent* discrete variables:



Same function except for a scale factor



$$P\,r(x|y=y_1) = P\,r(x,y_1)\,/P\,r(y_1)$$



$$P\,r(x|y=y_2) = P\,r(x,y_2)\,/P\,r(y_2)$$

equal

Equal, because x and y independent

$$Pr(x|y) = Pr(x)$$

Marginalization

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

(Courtesy: Prince, 2012)

Independence

$$Pr(x|y) = Pr(x)$$
$$Pr(y|x) = Pr(y)$$

Remember:

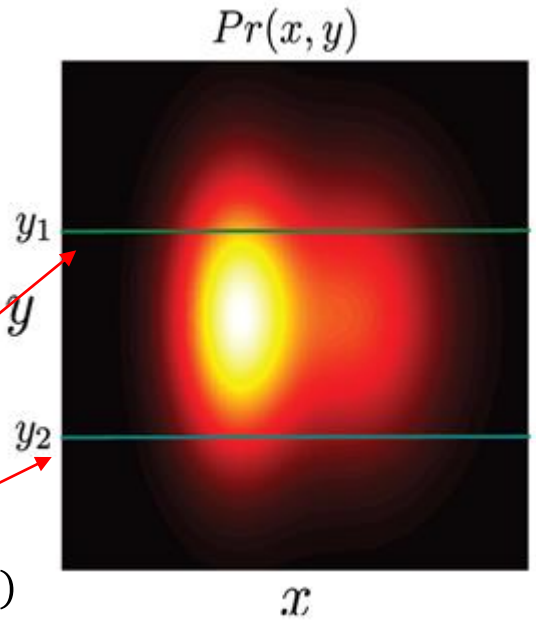
$$\int P(x, y_1)dx = P(y_1)$$

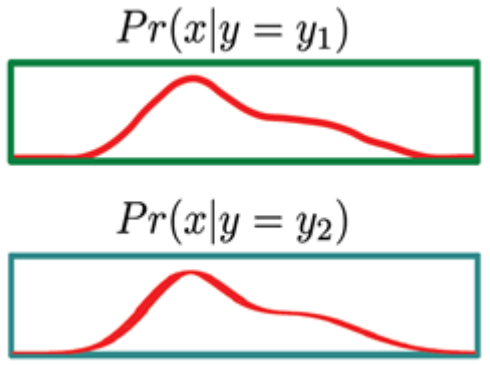
$$\int P(x|y_1)dx = 1$$

Example for *x*, *y independent* continuous variables:

$$\int P(x, y_1)dx$$
$$= \int P(x|y_1)P(y_1)dx$$
$$= P(y_1)\int P(x|y_1)dx$$

$= 1$



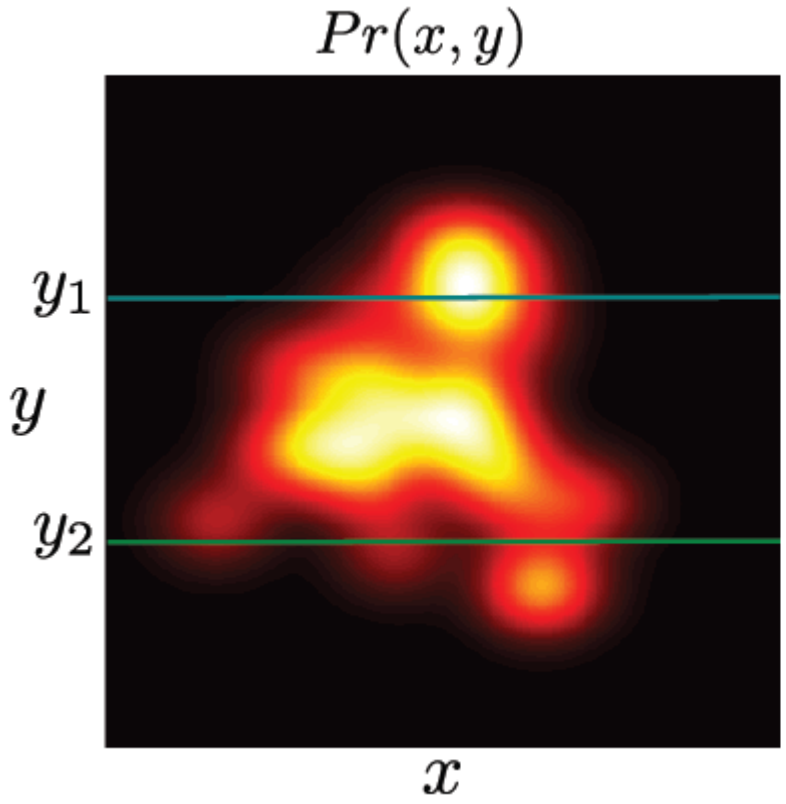


Same function except for a scale factor

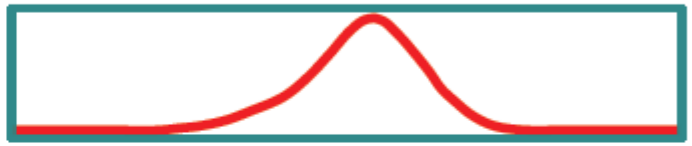
(Courtesy: Prince, 2012)

Not Independence (dependence)

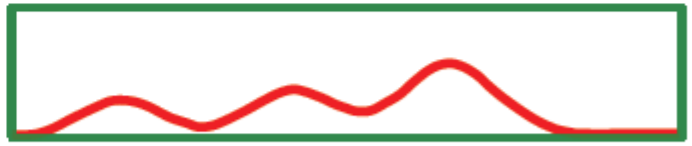
$Pr(x|y)$ depends on the especific value of y



Different for each value of y



$Pr(x|y = y_1)$



$Pr(x|y = y_2)$

x and y are **not independent** (are dependent)

(Courtesy: Prince, 2012)

Recapitulating: Two important rules

Sum Rule (Marginalization): $P(X) = \sum_{Y} P(X, Y)$

Product Rule: $P(X, Y) = P(Y \mid X) P(X)$

Law of total probability: $P(X) = \sum_{Y} P(X|Y) P(Y)$

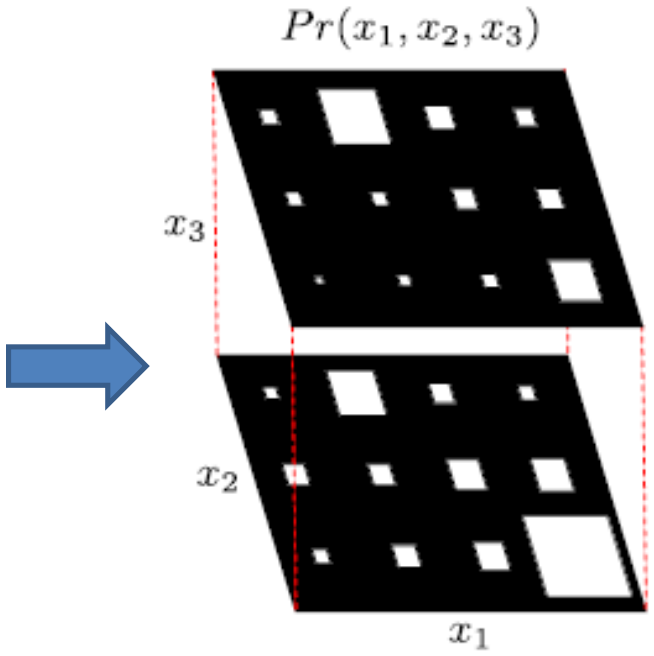
[Same applies for continuous RV]

Joint probability of 2+ RVs

Probability of 3 RVs: $P(x_1, x_2, x_3)$

Function of 3 variables→ lot of data needed

Example: x_1=1:4, x_2=1:3, x_3=1:2 → 24 entries



Different factorizations are possible:

$$P(\overbrace{x_1, x_2}^{x}, \overbrace{x_3}^{y}) = P(\overbrace{x_1, x_2}^{x} \mid \overbrace{x_3}^{y}) P(\overbrace{x_3}^{y}) = P(x_2|x_1, x_3)P(x_1|x_3)P(x_3)$$

$$= P(x_1|x_2, x_3)P(x_2|x_3)P(x_3)$$

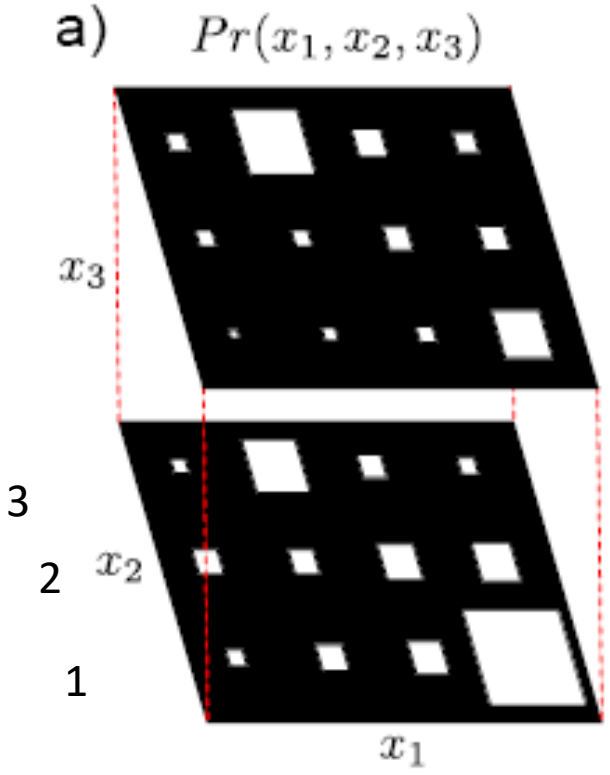
$$= P(x_3|x_1, x_2)P(x_1, x_2) = P(x_3|x_1, x_2)P(x_1|x_2)P(x_2)$$

$$= \cdots$$ More possibilities exist

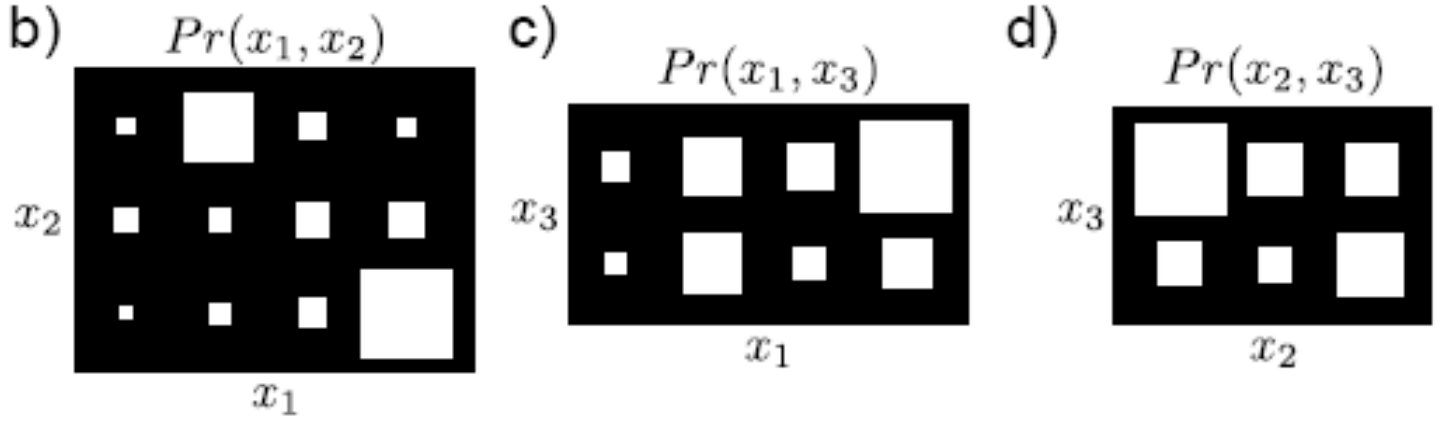
Given $P(x_1, x_2, x_3)$ we can compute all the *marginals* and *conditionals*

Example: Three discrete RVs (x_1=1:4, x_2=1:3, x_3=1:2)

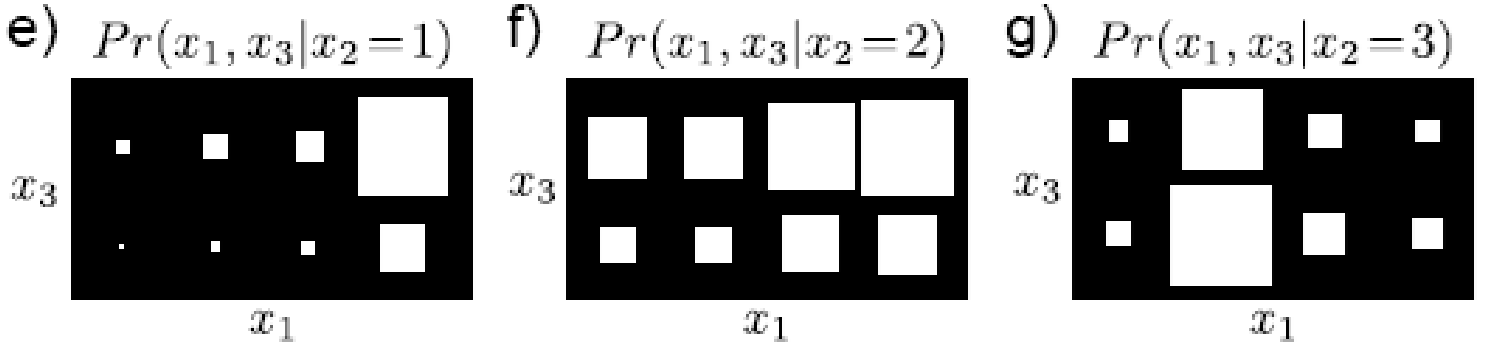
Joint of the 3 variables



Marginals



Conditionals



(Courtesy: Prince, 2012)

Conditioning to a third RV

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

Marginalizing out y

If we introduce a **third random variable (RV)** z:

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

Marginalizing out y

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

Conditional Independence (CI)

At least 3 RVs in play!

x conditionally independent of y (and viceversa) **given** $\boldsymbol{z}$ *if*

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

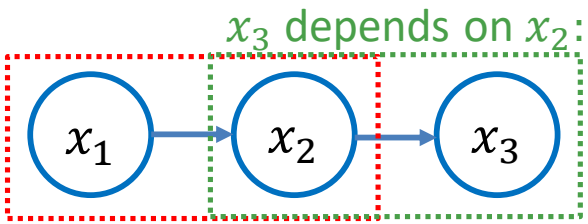
That is: $P(x \mid z) = P(x \mid z, y)$ and $P(y \mid z) = P(y \mid z, x)$

Be aware of:

- CI does not imply I: $P(x, y \mid z) = P(x \mid z)P(y \mid z) \not\Rightarrow P(x, y) = P(x)P(y)$
- I does not imply CI: $P(x, y) = P(x)P(y) \not\Rightarrow P(x, y \mid z) = P(x \mid z)P(y \mid z)$

Graphical Models

Bayesian network (aka DAG: directed acyclic graph → no loops):
A convenient way of representing the dependencies between variables

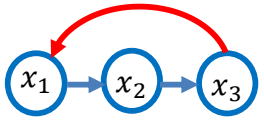


$$P(x_1, x_2, x_3) = P(x_3|x_2)P(x_2|x_1)P(x_1)$$

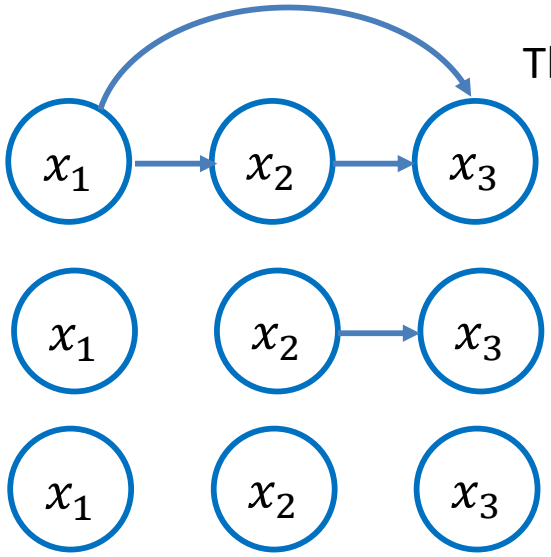
x_1 CI of x_3 *given* x_2: $P(x_3|x_1, x_2) = P(x_3|x_2)$

A **Bayesian network** tells us how to factorize a joint distribution:

This IS a cycle



This is not a cycle



$$P(x_1, x_2, x_3) = P(x_3|x_1, x_2)P(x_2|x_1)P(x_1)$$

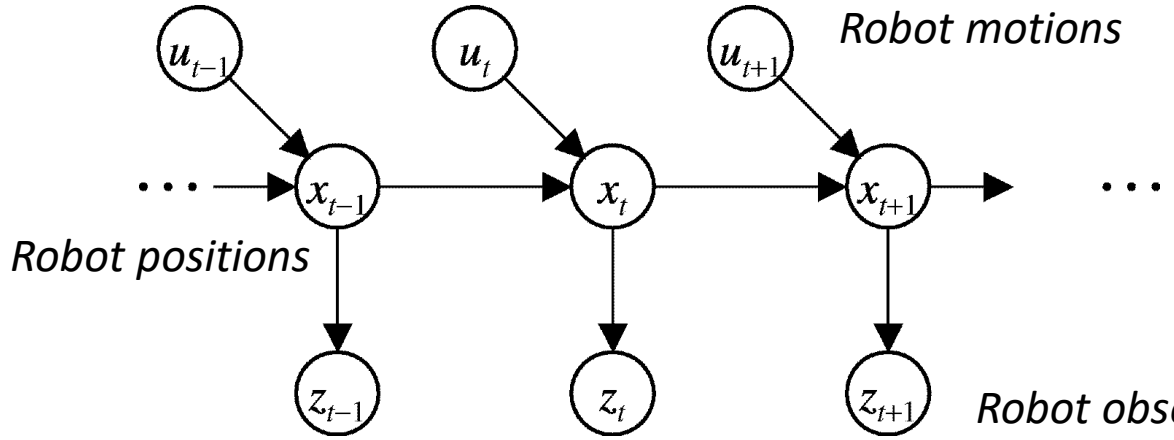
$$P(x_1, x_2, x_3) = P(x_3|x_2)\,P(x_2)P(x_1)$$

$$P(x_1, x_2, x_3) = P(x_3)P(x_2)P(x_1)$$

Markov assumption

For random variables that evolve over time: The probability of future states depends only upon the present state, not on the sequence of events that precededs it (i.e. the process is **memoryless**)

Example: Graphical model (Bayesian network) applying Markov assumption:



[Recall: If x and y are independent then $P(x|y) = P(x)$]

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

The robot position x at time t+1 does not depend on previous positions 1:t-1 **GIVEN the current robot position x_t**

$$p(z_{t+1}|x_{t+1}, z_{1:t}) = p(z_{t+1}|x_{t+1})$$

The observation z_{t+1} (at time t+1) does not depend on previous observations **GIVEN the current robot position x_{t+1}**

Expected value of a RV

- Average value of a random variable x
- Not necessarily the most probable value , which is the MODE

Definition:

$$E[x] = \sum_x x\,P(x) \qquad x \text{ discrete}$$

$$E[x] = \int x\,p(x)dx \qquad x \text{ continuous}$$

$P(x)/p(x)$ is the probability function of x

Notice: these are the definitions of the average value (mean) of the RV x

Example: The expected value of rolling a fair dice

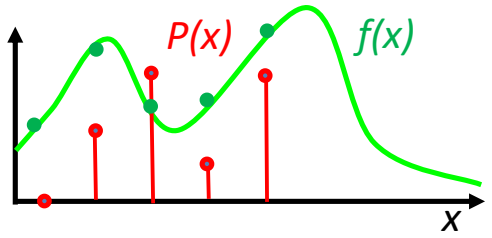
$$E[x] = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = 3.5$$

Expected value of a function

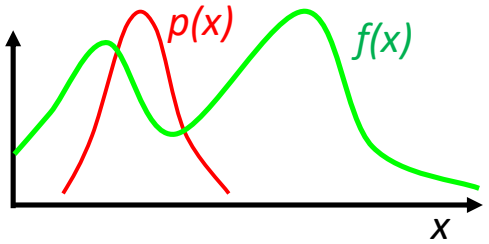
If x is a RV → $f(x)$ is also a RV (e.g. the square of the output of rolling a dice)

Expected value (or expectation) of a function $f(x)$, with x being a RV with distribution $p(x)$, is the average value of $f(x)$

x discrete $$E[f] = \sum_x P(x) f(x)$$



x continuous $$E[f] = \int p(x) f(x) dx$$



The expectation is always a number, not a RV!!

Expectation: Common Cases. Moments of a RV

Function $f[\bullet]$	Expectation
x	mean, μ_x
x^k	k^{th} moment about zero
$(x-\mu_x)^k$	k^{th} moment about the mean
$(x-\mu_x)^2$	variance
$(x-\mu_x)^3$	skew
$(x-\mu_x)^4$	kurtosis
$(x-\mu_x)(y-\mu_y)$	covariance of x and y

Expectation: Rules

Expectation is a linear function

$$\mathrm{E}[\kappa] = \kappa$$

$$\mathrm{E}[kf[x]] = k\mathrm{E}[f[x]]$$

$$\mathrm{E}[f[x]+g[x]] = \mathrm{E}[f[x]]+\mathrm{E}[g[x]]$$

$$\mathrm{E}[f[x]g[y]] = \mathrm{E}[f[x]]\mathrm{E}[g[y]] \qquad \text{if } x, y \text{ independent}$$

Variance and **Covariance**

- x is a scalar RV ($x \in \mathbb{R}$):

$$\underbrace{\sigma^2 = \mathrm{var}[x] = \mathrm{E}[(x-\mu)^2]}_{\text{definition}} = \mathrm{E}[x^2 - 2x\mu + \mu^2] = \mathrm{E}[x^2] - 2\mu^2 + \mu^2 = \mathrm{E}[x^2] - \mu^2$$

- $\boldsymbol{x} = [x_1, \ldots, x_n]^T$ is a vector of RVs ($\boldsymbol{x} \in \mathbb{R}^n$):

Mean vector $\mathrm{E}[\mathbf{x}]$

$$\underbrace{\mathrm{var}[\boldsymbol{x}] = \mathrm{E}[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T] = \mathrm{E}[\boldsymbol{x}\boldsymbol{x}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T}_{\text{definition}} = \Sigma$$

Σ is a nxn matrix called covariance matrix

$$\Sigma = [\sigma_{ij}{}^2] \text{ with } \sigma_{ij}{}^2 = E[x_i x_j] - \mu_i \mu_j = E[x_i x_j] - E[x_i]E[x_j]$$

If x_i and x_j are uncorrelated: $\sigma_{ij}{}^2 = 0 = E[x_i x_j] - \mu_i \mu_j \rightarrow E[x_i x_j] = E[x_i]E[x_j]$

Variance and **Covariance**

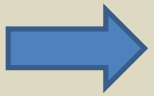
Example 2D:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad E[\mathbf{x}] = \boldsymbol{\mu} = \begin{bmatrix} E[x_1] = \mu_1 \\ E[x_2] = \mu_2 \end{bmatrix}$$

$$x - E[x] = \begin{bmatrix} x_1 - E[x_1] \\ x_2 - E[x_2] \end{bmatrix} = \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} E[\Delta x_1{}^2] & E[\Delta x_1 \Delta x_2] \\ E[\Delta x_1 \Delta x_2] & E[\Delta x_2{}^2] \end{bmatrix} = \begin{bmatrix} E[x_1{}^2] & E[x_1 x_2] \\ E[x_1 x_2] & E[x_2{}^2] \end{bmatrix} - \begin{bmatrix} \mu_1{}^2 & \mu_1 \mu_2 \\ \mu_1 \mu_2 & \mu_2{}^2 \end{bmatrix}$$

x_i x_j independent:
$p(x_i,x_j) = p(x_i)\,p(x_j)$



x_i x_j are uncorrelated!
$E[x_i x_j] = E[x_i]E[x_j]$

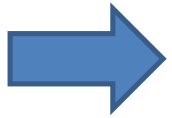
Proof:

$$E[x_i x_j] = \int\int x_i x_j\, p(x_i x_j)\,dx_i dx_j = \int\int x_i x_j\, p(x_i)\,p(x_j)\,dx_i dx_j =$$
$$= \int x_i\, p(x_i)\,dx_i \int x_j\, p(x_j)\,dx_j = E[x_i]E[x_j]$$

In general :

Uncorrelation  Independence

BUT, if the **RV are Gaussians**

Uncorrelation  Independence

Bayes' Theorem

$$P(X = x|Z = z) = \frac{P(Z = z|X = x)P(X = x)}{P(Z = z)} = p(x)\frac{p(z|x)}{\sum_x p(z|x)\,p(x)}$$

($p(z|x)$ over $P(Z = z|X = x)$; $p(x)$ over $P(X = x)$)

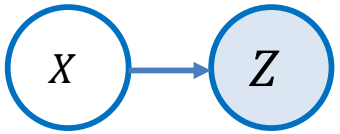
posterior $\propto$ likelihood × prior

Law of total probability

Tells us how a probability $p(x)$ changes when having an evidence z (observation)

$p(x|z)$ can be either > or < $p(x)$ depending on how 'good' the evidence z is

Bayes allows us to do INFERENCE:



Z depends on X, but we want to INFER the probability of X given Z

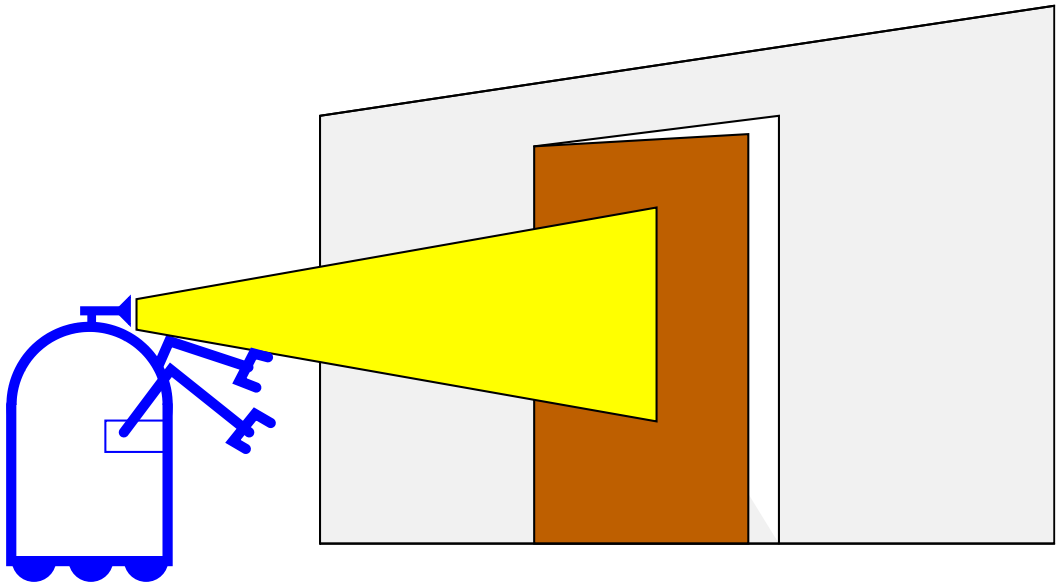
Shaded means GIVEN/OBSERVED

[We'll see an example of this next]

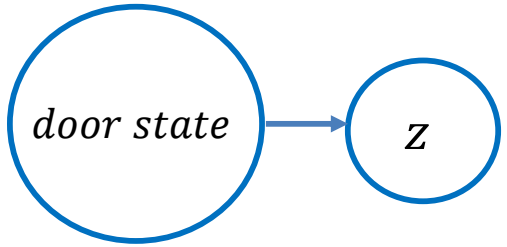
Simple Example of State Estimation

- Suppose a robot has an **observation *z* that a door is open: $z^{o(pen)}$**
- What is the probability that the door is *open* if (conditioned to) the sensor tells us that it is *open* (z^o)?

$$P(\text{door state} = \text{open} \mid z^o)$$



Graphical Model



Both, $door\ state$ and z are Boolean variables

$$door\ state \in \{open, \neg open\}$$

$$z \in \{z^0, \neg z^0\}$$

Simple Example of State Estimation

Known information

- $P(z^o|open) = 0.6$ (True Positive rate)
  $P(z^o|\neg open) = 0.3$ (False Positive rate)
- $P(open) = P(\neg open) = 0.5$

Sensor model: how accurate the sensor is

Prior Information: How much we know about the world

Please, notice: $P(z^o|open) + P(z^o|\neg open) \neq 1$

$$P(z^o|open) + P(\neg z^o|open) = 1$$

Applying Bayes:

$$P(open|z^o) = \frac{P(z^o|open)P(open)}{P(z^o|open)p(open) + P(z^o|\neg open)p(\neg open)} = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{2}{3} = 0.67$$

The observation z^o raises the probability that the door is open (0.5 → 0.67)

Combining Evidences

Suppose our robot gets another observation z_2^o (z_2, for simplicity)

– How can we integrate this new information?

– Specifically, how can we estimate $P(x|z_1, z_2)$?

Answer: Recursive Bayesian Updating

2 observations:

Bayes on z_2: all probabilities
conditioned on the first observation z_1

$$P(x|z_1, z_2) = \frac{P(z_2|x, z_1)\ P(x|z_1)}{P(z_2|z_1)}$$

Example: Second Measurement

- $P(z_2|open) = 0.5 \quad P(z_2|\neg open) = 0.6$ ← Not a good sensor!
- $P(open|z_1)=2/3$ ← New prior

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

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

z_2 lowers the probability that the door is open (0.67 → 0.625)

Notice that if $P(z_2|open) = P(z_2|\neg open) = 0.5$ the new observation gives no information (0.67 → 0.67)

Recursive Bayesian Update

n observations

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

Bayes rule for the last observation z_n

Markov

z_n *does not depend on previous z given x*

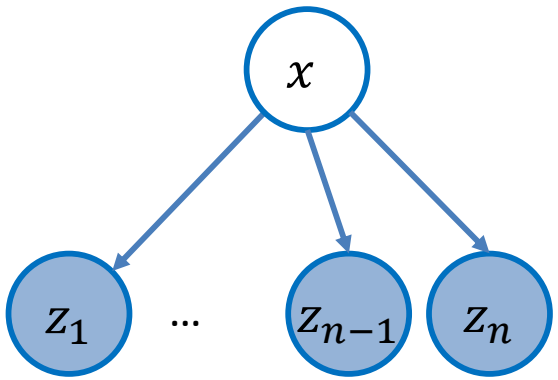
$$= \frac{P(z_n|x)\, P(x|z_1, \dots, z_{n-1})}{P(z_n|z_1, \dots, z_{n-1})}$$

$P(z_n|z_1, \dots, z_{i-n}) = 1/\eta_n$

$$P(x|z_1, \dots, z_n) = \eta_n\, P(z_n|x)\, P(x|z_1, \dots, z_{n-1})$$

P updated after a new observation z_n

Bayesian network



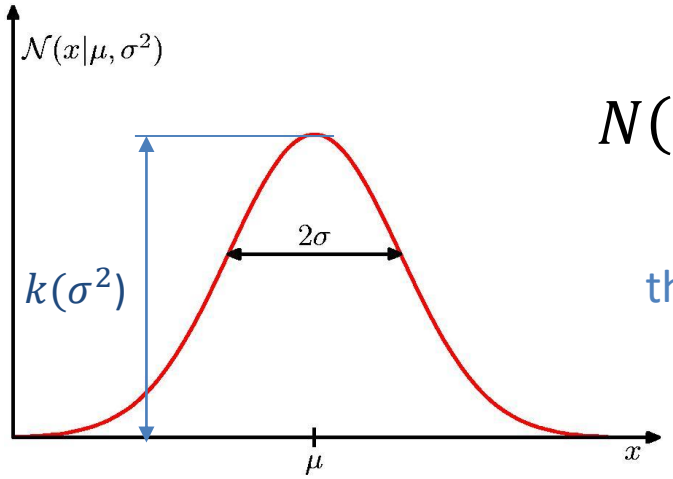
$$P(x|z_1, \dots, z_n) = P(x) \prod_{i=1\dots n} \eta_i\, P(z_i|x)$$

P using all the observations at once [Applying Recursivity]

$$P(x|z_1, \dots, z_n) \propto \prod_{i=1\dots n} P(z_i|x)$$

If $P(x)$ is uniform (i.e. constant] $\eta_i = \eta$ constant

The Gaussian Distribution: 1D



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

the RV

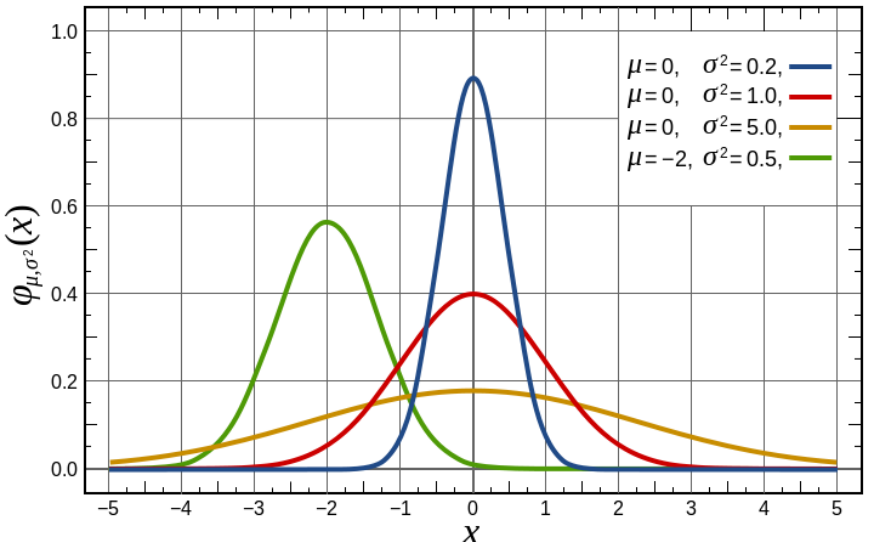
The parameters of the Gaussian

- At $x = \mu$ we have the MODE (highest density)

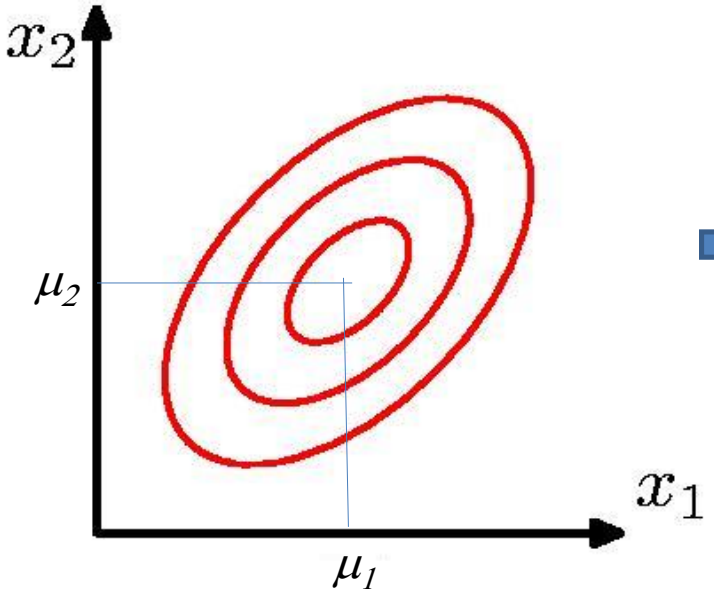
$$N(x=\mu;\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} = k(\sigma^2)$$ constant that depends on σ^2

- σ gives us the height of the bell, i.e., its width

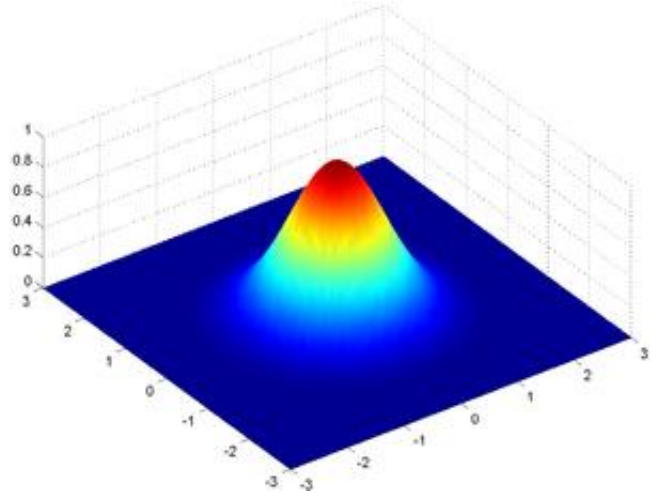
Recall: the area under the bell must be 1: $\int_{-\infty}^{\infty} N(x;\mu,\sigma^2)dx = 1$



The Gaussian Distribution: 2D



In 3D



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

Also denoted

$$\mathcal{N}(x;\mu,\Sigma)$$

A number that gives the **scale**

A quadratic polynomial that gives the **position** (μ) and the **shape** (Σ) of the bell

Gaussian Mean and Variance

$$\text{mean}[x] = \text{E}[x] = \int_{-\infty}^{\infty} N(x; \mu, \sigma^2) x dx = \mu$$

$$\text{var}[x] = \text{E}[(x - \text{E}(x))^2] = \text{E}[x^2] - \text{E}\big[2x\text{E}[x]\big] + \text{E}[\text{E}[x]^2]$$

$$= \text{E}[x^2] - \text{E}[x]^2 = \text{E}[x^2] - \mu^2 = \sigma^2$$

definitions

$$\text{E}[x^2] = \int_{-\infty}^{\infty} N(x; \mu, \sigma^2) x^2 dx = \mu^2 + \sigma^2$$

$$\text{mean}[x] = \mu$$
$$\text{var}[x] = \sigma^2$$

Properties of gaussian distributions

These two results are properties of Gaussians, not a definition

Geometry of the Multivariate Gaussian

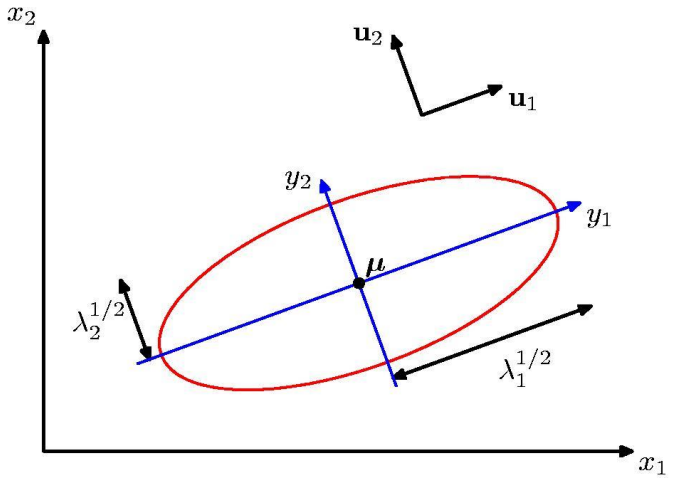
- Since Σ is a *symmetric semidefinite positive matrix*, the quadratic polynomial

$$\Delta(x)^2 = (x-\mu)^T \sum^{-1} (x-\mu)$$

$\Delta(\mathrm{x})$ is called Mahalanobis distance.

is the expression of an ellipse centered at μ and with shape given by $\sum$

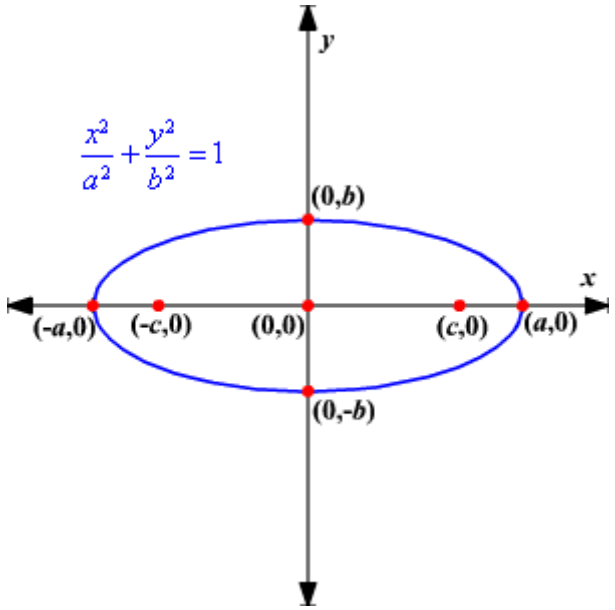
- The axes of the ellipse are defined by the two eigenvectors u_1, u_2 of the covariance matrix (assuming $x \in \mathbb{R}^2, \sum \in \mathbb{R}^{2x2}$)



Geometry of the Multivariate Gaussian

EXAMPLE 1

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \Rightarrow \Sigma^{-1} = \frac{1}{6}\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$



$$\Delta(x)^2 = (x-\mu)^T\Sigma^{-1}(x-\mu) = \frac{1}{6}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{{x_1}^2}{3} + \frac{{x_2}^2}{2}$$

All the points x on the ellipse are at equal distance $\Delta(x)^2$ to μ

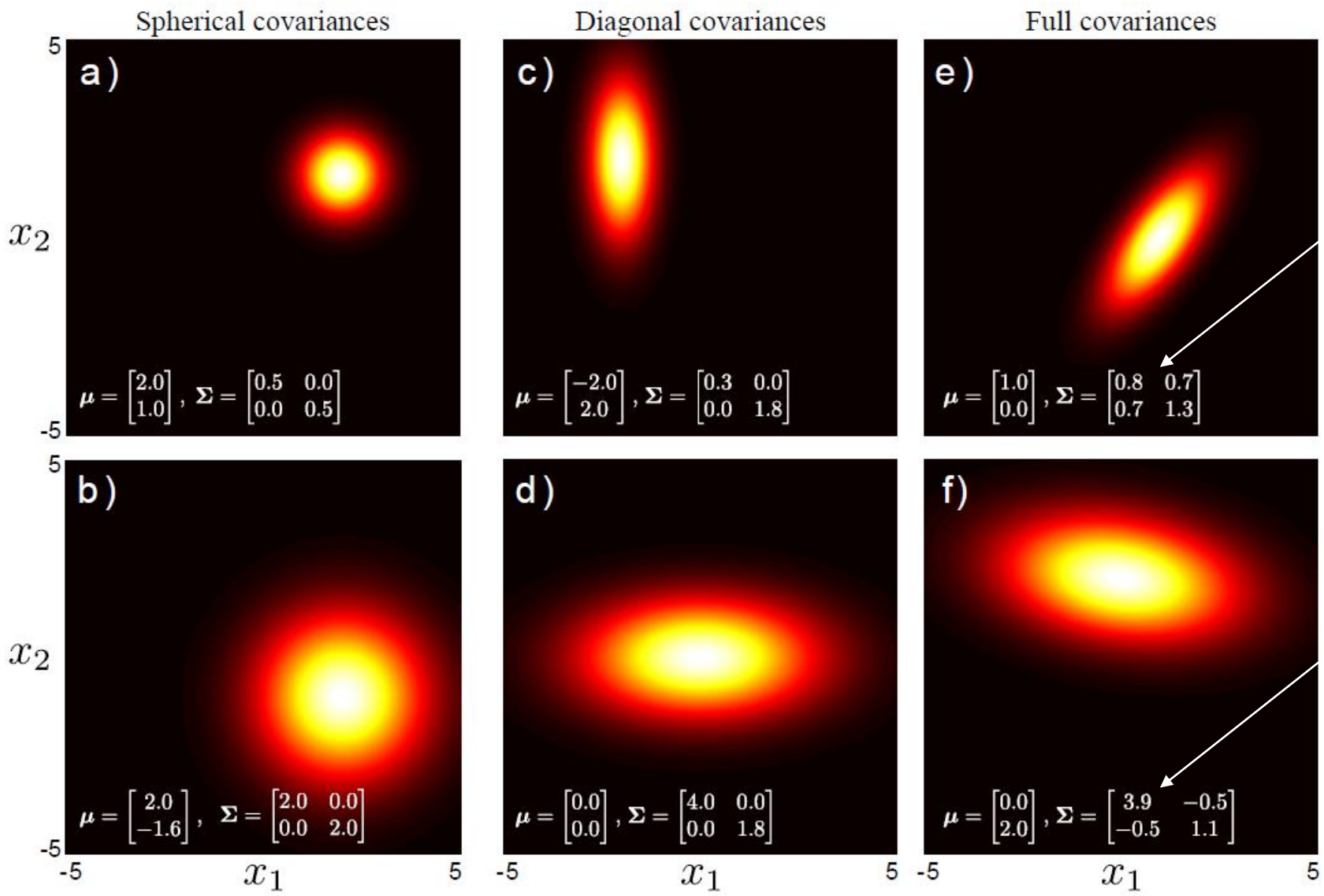
EXAMPLE 2

If Σ is not diagonal → Σ^{-1} is not diagonal and the ellipse is not aligned to the (x_1, x_2) axis.

$$\Sigma^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \quad \Delta(x)^2 = (x-\mu)^T\Sigma^{-1}(x-\mu) = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2{x_1}^2 + 3{x_2}^2 + 2x_1x_2$$

If the mean $\mu \neq 0$ the ellipse is centered at μ

Geometry of the Multivariate Gaussian



Positive correlation

When
$\Delta x = (x - \mu_x) > 0 \rightarrow \Delta y = (y - \mu_y) > 0$

Positive correlation

When $\Delta x > 0 \rightarrow \Delta y < 0$

The Covariance matrix Σ tells us about the shape and size of the ellipse

Conditionals and Marginals for Gaussians

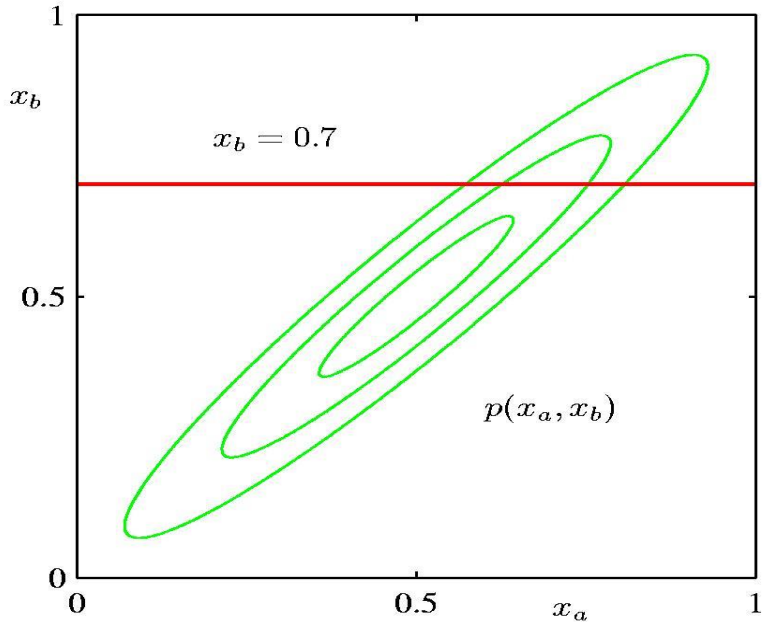
We have two o more Gaussian RVs: x_a, x_b, …

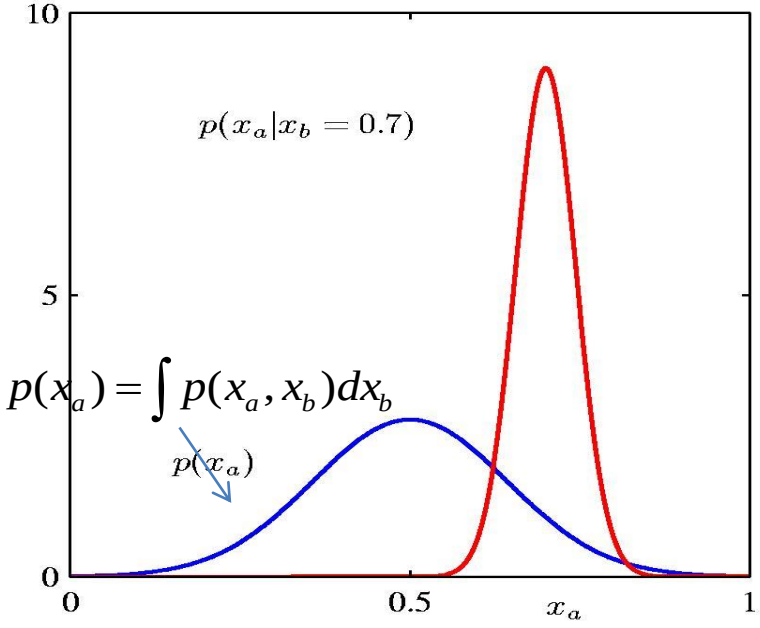
For example: robot position coordinates x, y

Notice that for different values of x_b the shape of $p(x_a|x_b)$ es different

x_a and x_b are not independent:

$$p(x_a) \neq p(x_a|x_b)$$





If the Gaussian is aligned to the x_a, x_b axes → they are independent → Not correlated (Diagonal covariance matrix)

On the covariance matrix Σ

More about Σ :

Σ (and Σ^{-1}) is symmetric positive definite matrix

- The analogous of positive number
- Definition: $M \succ 0$ iff $z^T M z > 0 \quad \forall z \neq 0$

Example: $M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = z_1^2 + z_2^2 > 0$

Properties:

- Invertible (with symmetric positive definite inverse)
- **Trace** (sum of the diagonal entries) and **determinant** are > 0
- Can be decomposed as $M = LL^T$ (L is upper triangular) by **Cholesky** decomposition
  (quite important for efficient of linear equations $Mx = b \rightarrow L(L^T x) = b$)

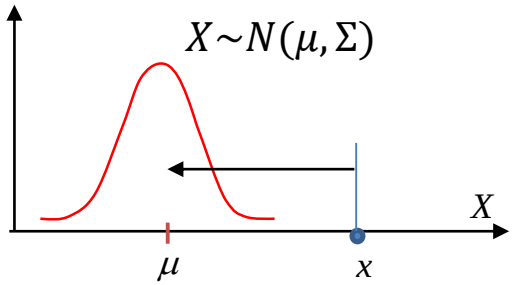
Euclidean vs. Mahalanobis distance

Euclidean distance from x to μ: $d^2(x,\mu) = (x-\mu)^T(x-\mu)$

Mahalanobis distance from a data point x to a distribution $N(\mu,\Sigma)$

if $X \sim N(\mu,\Sigma)$: $MD^2_{\mu,\Sigma}(x) = (x-\mu)^T\Sigma^{-1}(x-\mu)$

1D

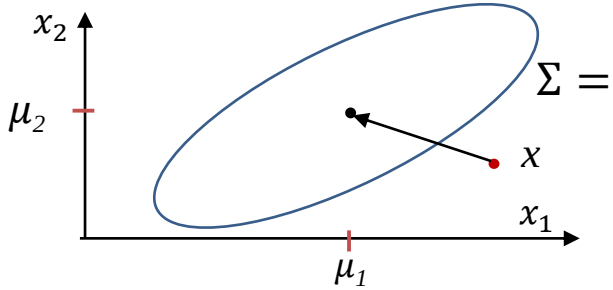


$$d^2(x,\mu) = (x-\mu)^2$$

$$MD^2_{\mu,\Sigma}(x) = \left(\frac{x-\mu}{\sigma}\right)^2$$

It's like the Euclidean distance but scaled (dimensional-less)

2D



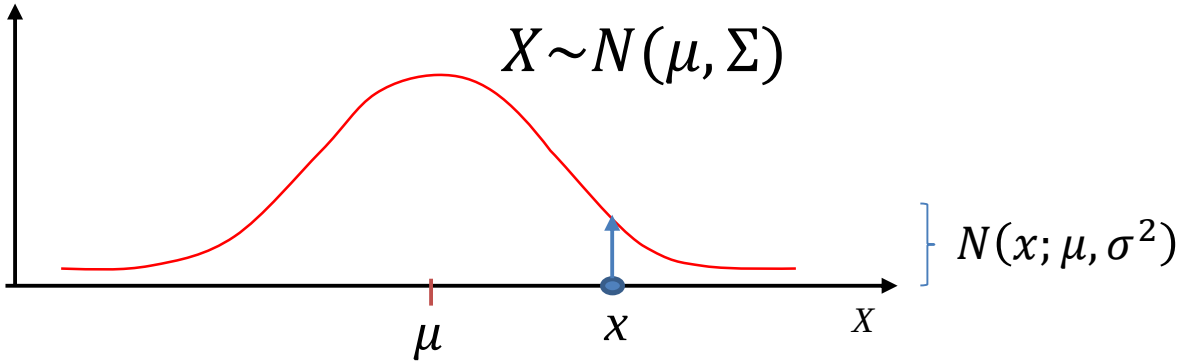
$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{bmatrix}$$

σ_{12} can not be arbitrary, must fulfill $|\Sigma| > 0$

Tells us how far is a data point to a normal distribution.

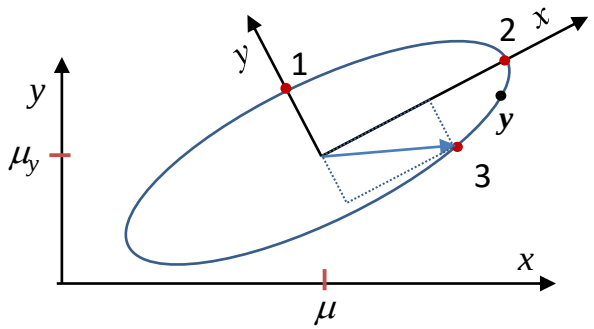
The closest a sample can be from a RV is when it is at the mean μ (*MD*=0)

Mahalanobis distance and probability density



The height of the gaussian probability density at x is given by the $MD^2_{\mu,\Sigma}(x)$ (up to a scale factor depending on σ)

$$N(x;\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \qquad = K\exp\left\{-\frac{1}{2}MD(x)^2\right\}$$



Recall:

$\Delta(x)^2 = (x-\mu)^T \sum^{-1} (x-\mu)$ is a quadratic polynomial

Loci of the points with equal square MD (Δ^2) is an ellipse

Points (1, 2, 3)
- are at the same Mahalanobis Distance to the center μ
- have the same likelihood to come from the distribution

Properties of Gaussian Distributions

Why Gaussian distributions are so appealing?

Two major reasons

- Product
- Marginalization/conditioning
- Affine transformation

} of Gaussians pdf's is Gaussian

Central Limit Theorem

The distribution of the sum of N i.i.d. random variables becomes increasingly Gaussian as N grows

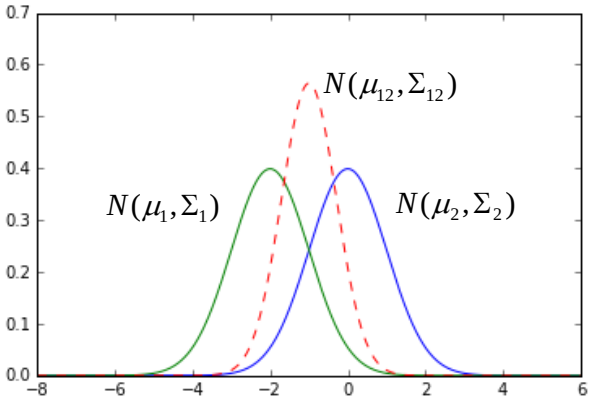
i.i.d. → independent: $p(x_i, x_j) = p(x_i)p(x_j)$

and identically distributed $x_i \sim N(\mu, \sigma^2)$

Product of Gaussians pdf is Gaussian

Do not memorize, just understand!

Example: $p(x_1, x_2) = p(x_1)\, p(x_2)$



$$\left.\begin{aligned} X_1 &\sim N(\mu_1, \Sigma_1) \\ X_2 &\sim N(\mu_2, \Sigma_2) \end{aligned}\right\} \quad \Sigma_{12}^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1} \Rightarrow \boxed{\Sigma_{12} = \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)^{-1}}$$

$$\boxed{\mu_{12} = \Sigma_{12}\left(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2\right)}$$

1D:

$$\boxed{\mu_{12} = \frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}}$$

$$\frac{1}{\sigma_{12}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2\sigma_2^2} \rightarrow \boxed{\sigma_{12}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}}$$

Always smaller than σ_1 and σ_1

$$\sigma_{12}^2 = \frac{\sigma_2^2}{1+\sigma_2^2/\sigma_1^2} < \sigma_2^2$$

$$\sigma_{12}^2 = \frac{\sigma_1^2}{1+\sigma_1^2/\sigma_2^2} < \sigma_1^2$$

If $\sigma_1 = \sigma_2 = \sigma$ (e.g. same sensor, as in the figure above)

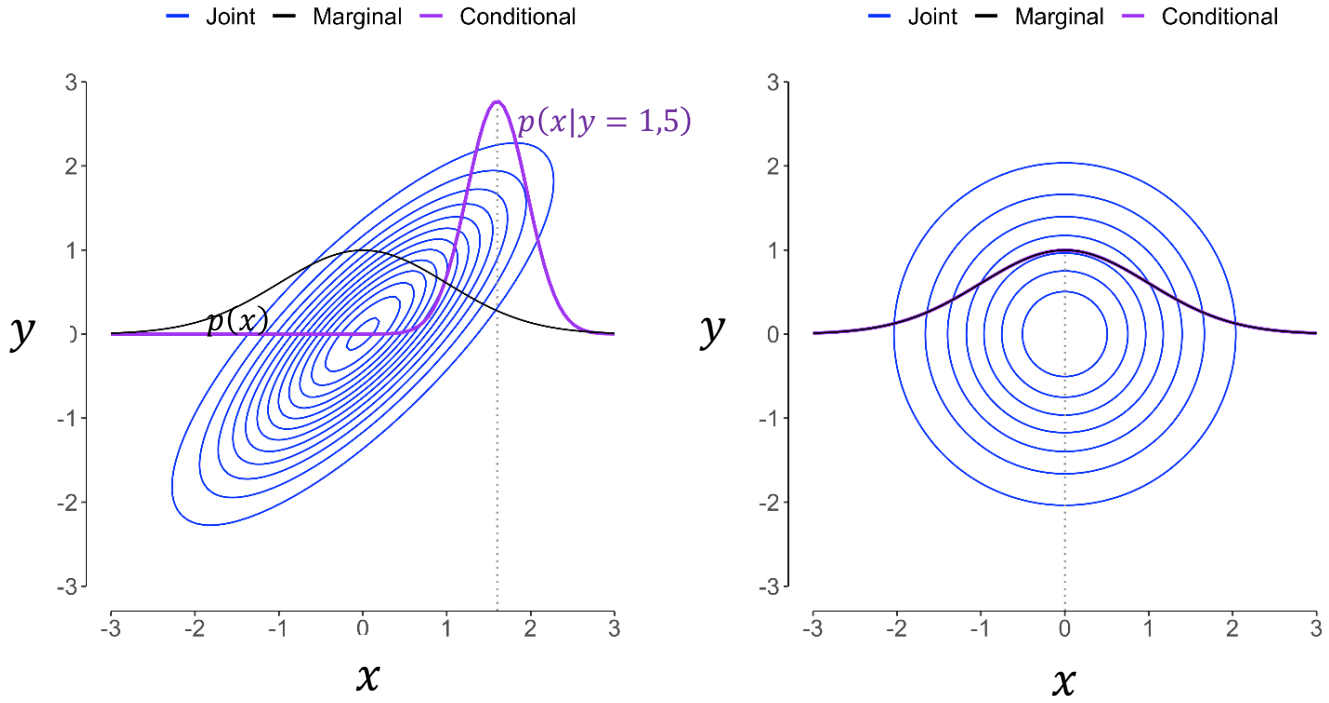
$$\mu_{12} = \frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2} = \frac{\mu_1 + \mu_2}{2} \qquad \sigma_{12}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma^2}{2}$$

Marginalization/conditioning of Gaussians pdf's is a Gaussian

$$p(x) = \int_y p(x,y)dy = \int_y p(x|y)p(y)\,dy$$

Do not memorize, just understand!

If $p(x,y) = \mathcal{N}\left(\begin{bmatrix}\mu_x \\ \mu_y\end{bmatrix}, \begin{bmatrix}\Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy}\end{bmatrix}\right)$ then



$$p(x) = \mathcal{N}(\mu_x, \Sigma_{xx})$$

$$p(y) = \mathcal{N}(\mu_y, \Sigma_{yy})$$

$$p(x|y) = \frac{p(x,y)}{p(y)} = \mathcal{N}\left(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)$$

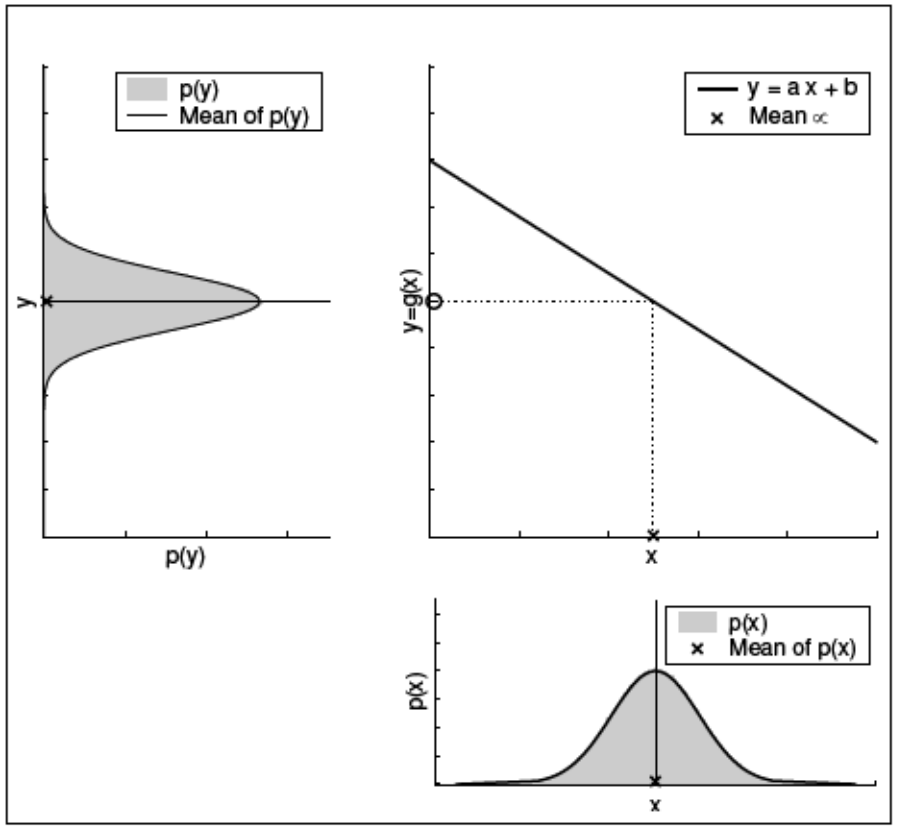
The mean changes with y, except if Σ diagonal ($\Sigma_{xy} = 0$)

Uncertainty propagation

Affine transformation of Gaussian is Gaussian

$$Y \sim N(A\bar{X} + B, A\Sigma_X A^T)$$

Another Gaussian



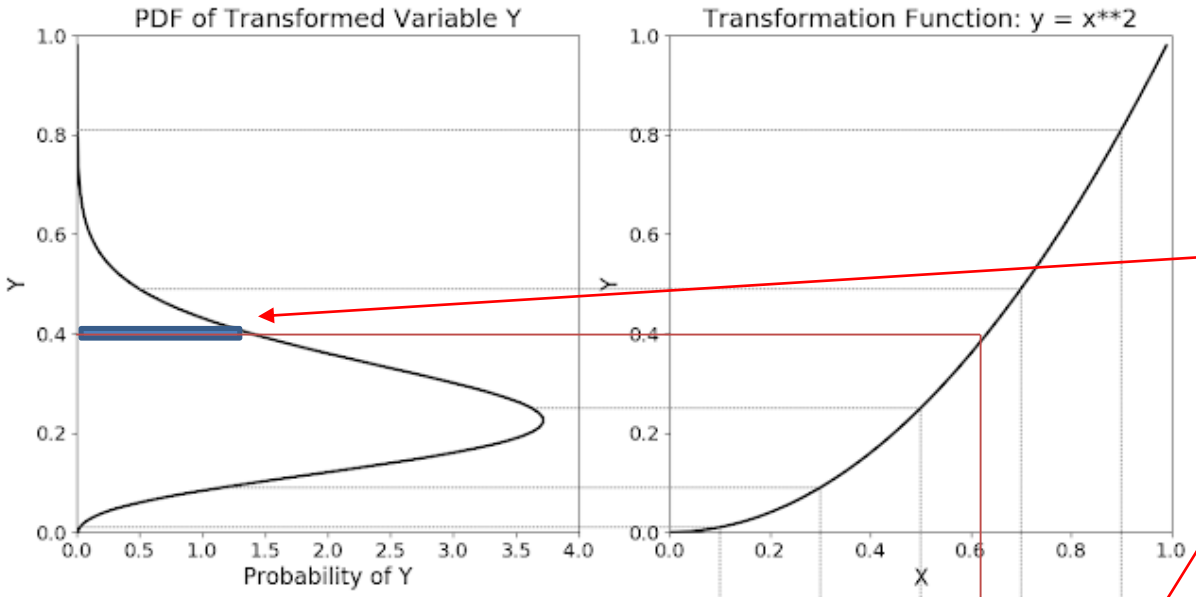
Affine transformation of RV X:

$$Y = AX + B$$

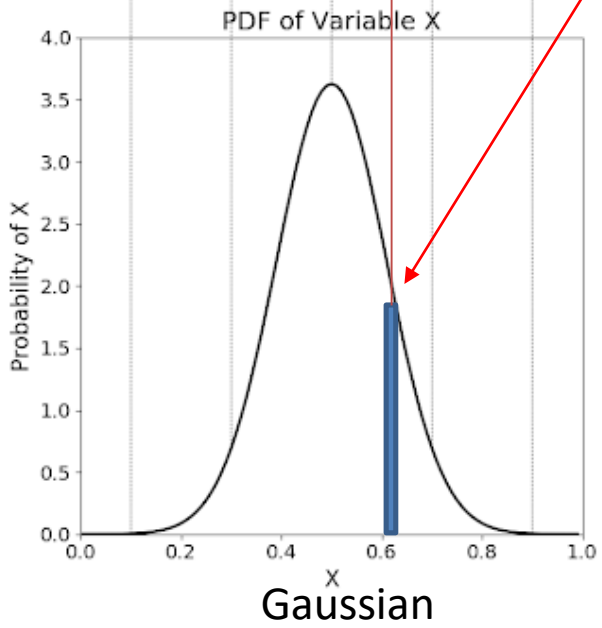
$$X \sim N(\bar{X}, \Sigma_X)$$

Depending on the slope A, the new RV can lower o raise the uncertainty (Σ)

What if the transformation between RVs is not linear?



Not Gaussian



Gaussian

Idea: The probability of x mas be the same that of $f(x)$

Two infinitesimal corresponding rectangles must have the same area (probability).

$$p_x(x)dx = p_y(y)dy$$

$$p_y(y) = p_x(x)\frac{dx}{dy} = p_x(x)\frac{1}{f'(x)}$$

Since $dy > dx$, the height $p_y(y)$ has to decrease

The transformed pdf is NOT Gaussian!

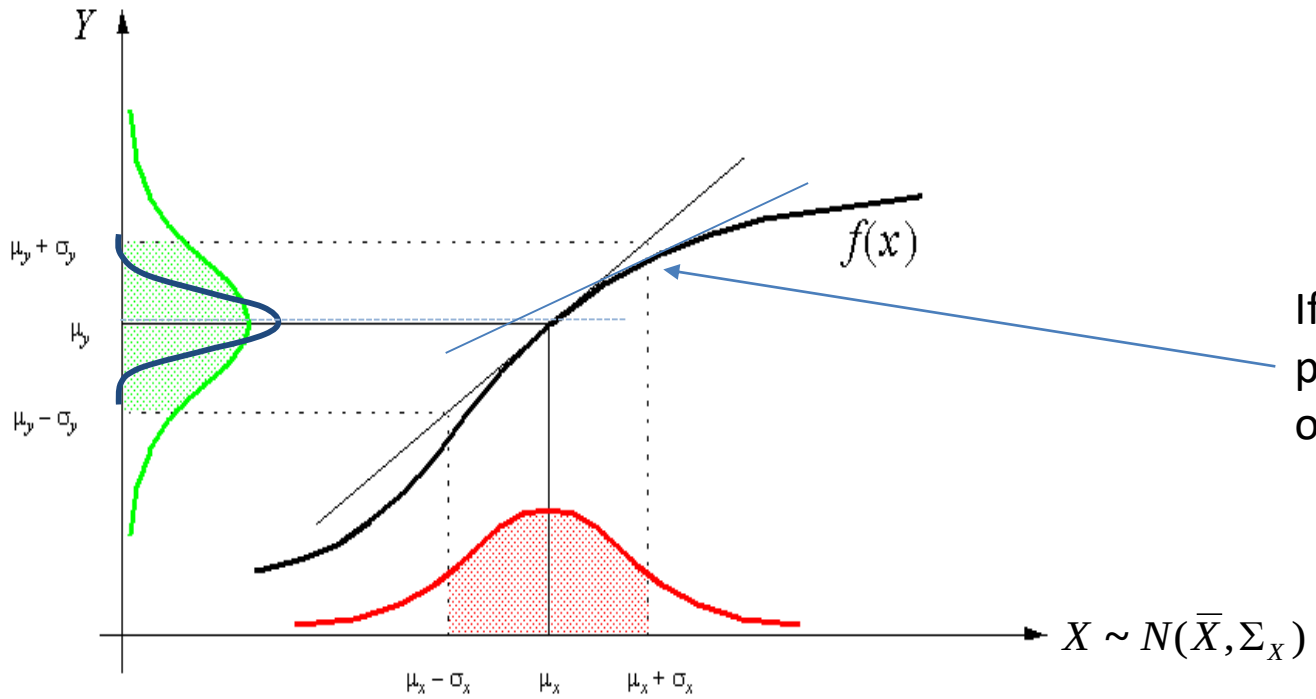
SOLUTION: Linearize at the mean $\bar{X}$ (Taylor approximation of a function)

$$Y = f(X) = f(\bar{X} + \Delta X) \approx \bar{Y} + \left.\frac{\partial f}{\partial X}\right|_{X=\bar{X}} \Delta X \quad \xrightarrow{\Delta Y = Y - \bar{Y}} \quad \Delta Y = \left.\frac{\partial f}{\partial X}\right|_{X=\bar{X}} \Delta X$$

$$f(\bar{X}) = \bar{Y} = \mathrm{E}[Y]$$

$$\Sigma_Y = E[\Delta Y \Delta Y^T] \approx E\left[\frac{\partial f}{\partial X} \Delta X \Delta X^T \frac{\partial f}{\partial X}^T\right] = \frac{\partial f}{\partial X} E[\Delta X \Delta X^T] \left(\frac{\partial f}{\partial X}\right)^T = \frac{\partial f}{\partial X} \Sigma_X \left(\frac{\partial f}{\partial X}\right)^T$$

Important: Jacobian evaluated at the mean $\bar{X}$



If we evaluate the Jacobian at a point other that the mean $\bar{X}$ we obtain a quite different Σ_Y

$$\Sigma_Y \approx \frac{\partial f}{\partial X} \Sigma_X \left(\frac{\partial f}{\partial X}\right)^T$$

Error propagation in the N-dimensional case:

$Y=f(X)$

$$\Sigma_Y = \frac{\partial f}{\partial X}\Sigma_X\left(\frac{\partial f}{\partial X}\right)^T$$

- Σ_X, Σ_Y: covariance matrices of X and Y
- $\frac{\partial f}{\partial X}$ is the **Jacobian** matrix defined as: $\frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f_1}{\partial X_1} & \cdots & \frac{\partial f_1}{\partial X_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial X_n} & \cdots & \frac{\partial f_m}{\partial X_n} \end{bmatrix}$

Dimension of X

Dimension of f

$Z=f(X,Y)$

$$\Sigma_Z = \underbrace{\frac{\partial f}{\partial X}\Sigma_X\left(\frac{\partial f}{\partial X}\right)^T}_{\Sigma'_X} + \underbrace{\frac{\partial f}{\partial Y}\Sigma_Y\left(\frac{\partial f}{\partial Y}\right)^T}_{\Sigma'_Y}$$

Still Covariance matrices, but projected in the space of Z

This will be used in a number of techniques in robotics:
Extended Kalman Filter, Least Square, …

Uncertainty (error) propagation in the N-dimensional case:

If $x_t = g(u_t , x_{t-1})$ is a linear function in both u_t and x_{t-1} :

$$x_t = A\, x_{t-1} + B + C\, u_t + D$$

then

$$\bar{x}_t = A\, \bar{x}_{t-1} + B + C\, \bar{u}_t + D$$

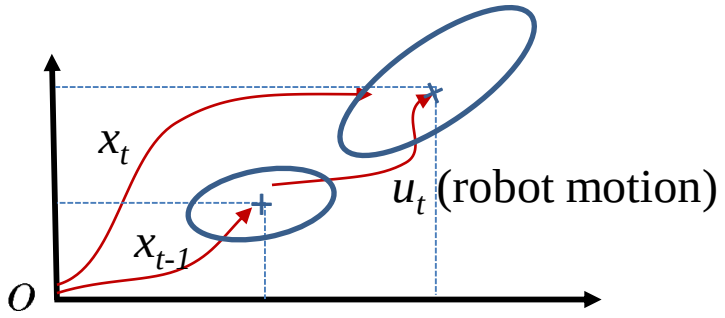
$$\Sigma_{x_t} = A\, \Sigma_{x_{t-1}} A^T + C\, \Sigma_{u_t} C^T$$

Example: Sum of random variables (e.g. pure robot translation)

$$x_t = A\, x_{t-1} + B + C\, u_t + D = x_{t-1} + u_t \qquad [A=C=1,\ B=D=0]$$

$$\bar{x}_t = \bar{x}_{t-1} + \bar{u}_t$$

$$\Sigma_{x_t} = \Sigma_{x_{t-1}} + \Sigma_{u_t}$$



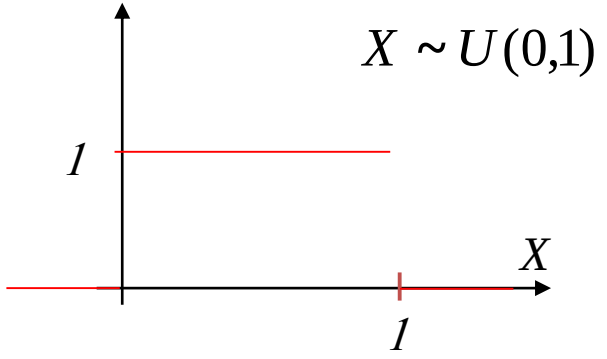
Some remarks

- Having the distribution (pdf) of a RV is the best we can aspire to (more complete than any statistics like the mode, the mean, …)
- With **Gaussians** is easy, we just need to compute the Mean and Covariance
- BUT, if any of the involved distributions is NOT Gaussian

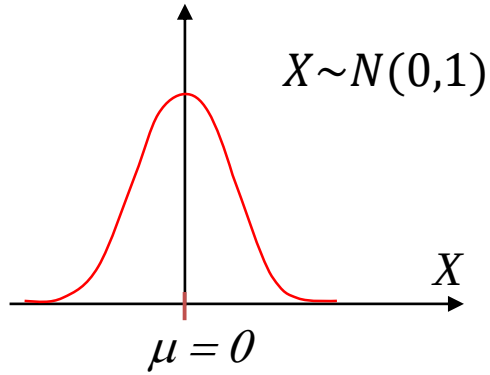
    1. Montecarlo approximations (Particle Filter)
    2. Alternatively, we can compute the best estimate:
        - the **most probable value** → the **MODE**
          Does not depend on the distribution
        - the **expected value** → the **MEAN**
          Takes into account the distribution

Sampling from a distribution

- Very useful for Montecarlo-based methods:
  - **Particle Filter** localization
  - Computing **Expectation** of a function
  - Knowing how **RVs transform** through a non-linear function
- Common assumption: RVs are *independent and identically distributed* (i.i.d.). Typically, samples taken from a given sensor
- Two basic distributions are mostly used:

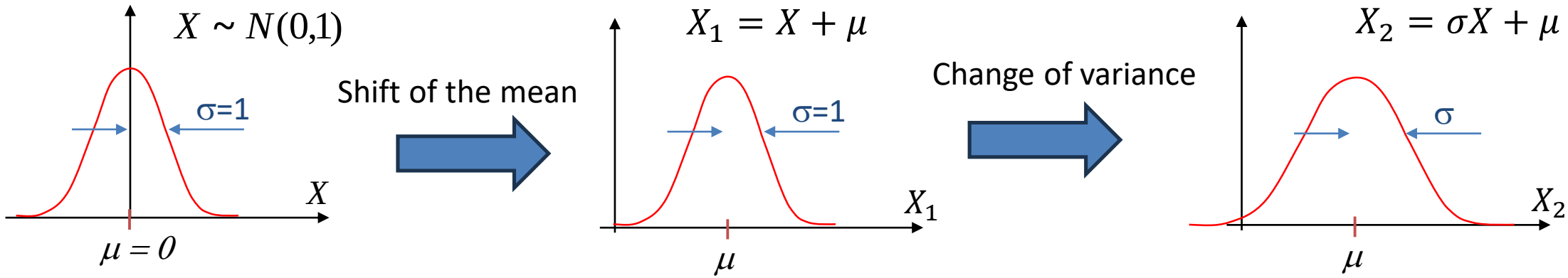


Standard Uniform: $X \sim U(0,1)$



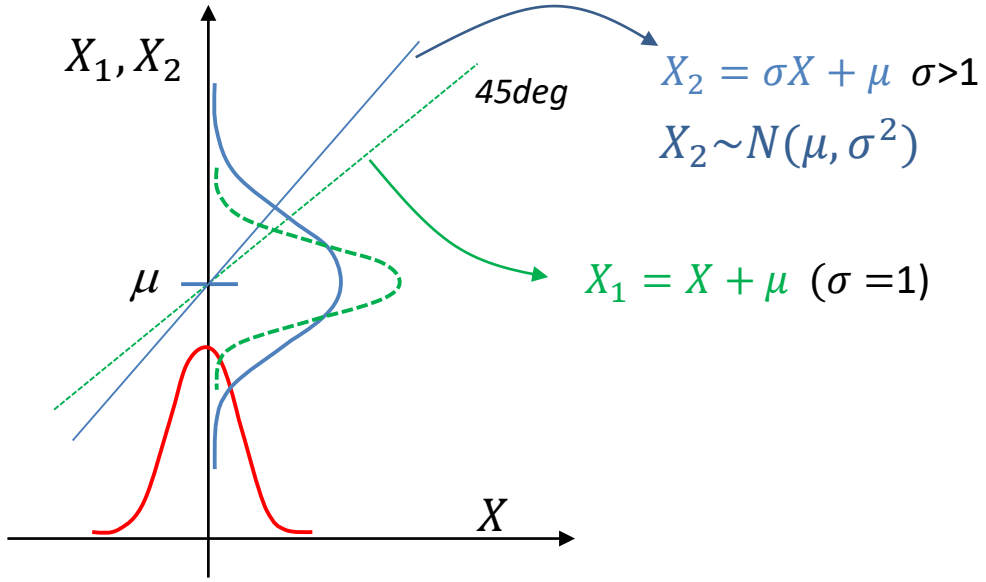
Standard Gaussian: $X \sim N(0,1)$

Non-Standard Gaussian: $X_1 \sim N(\mu, \sigma^2)$



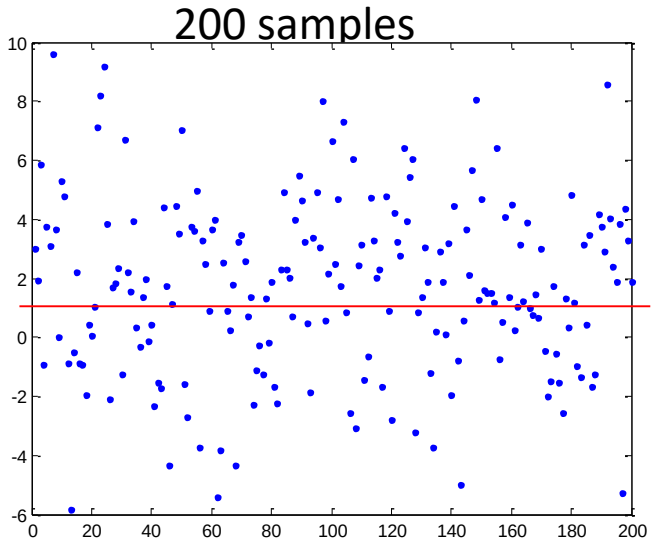
Standard Gaussian μ=0, σ=1

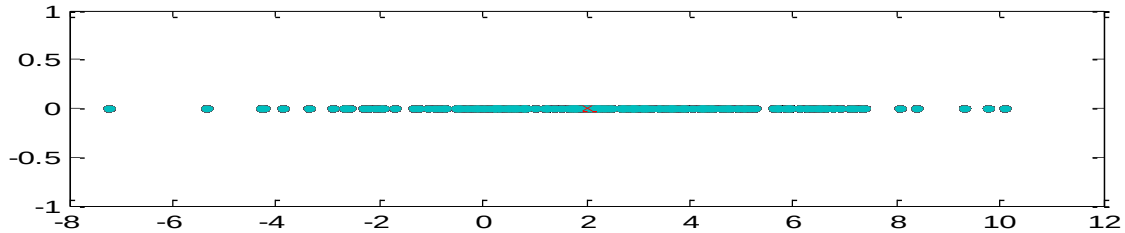
Recall:

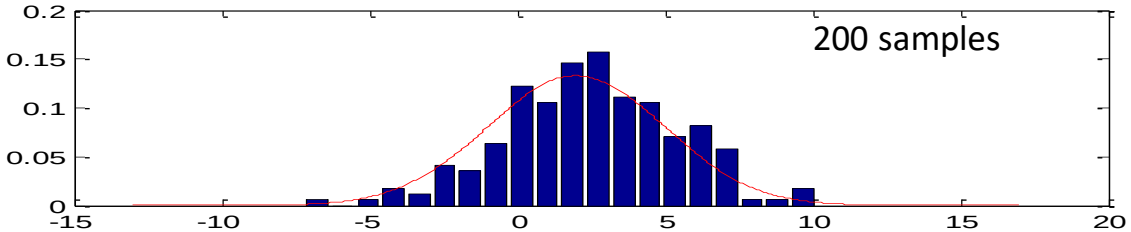


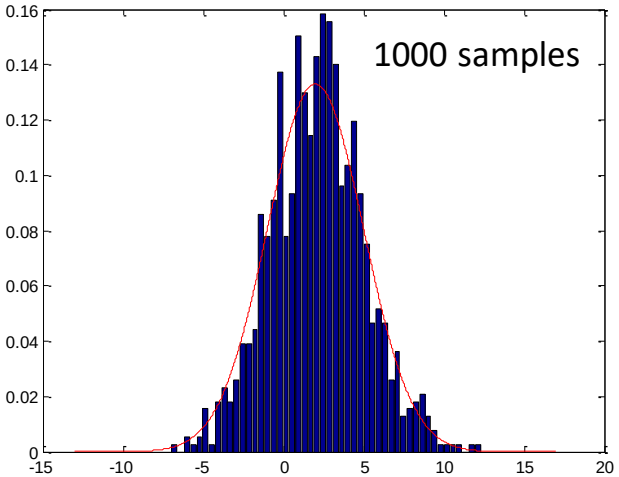

```
x=randn(samples,1); %vector of random numbers from a N(0,1)
x1 = mean + sigma*r; %Vector of sample for a non-standard gaussian
```

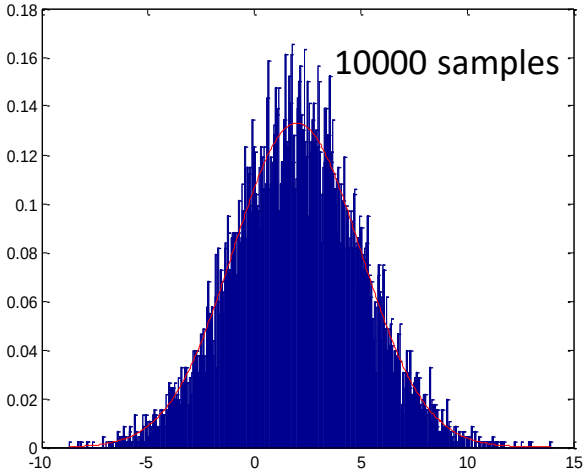
Example: Gaussian samples using N(1, 1)







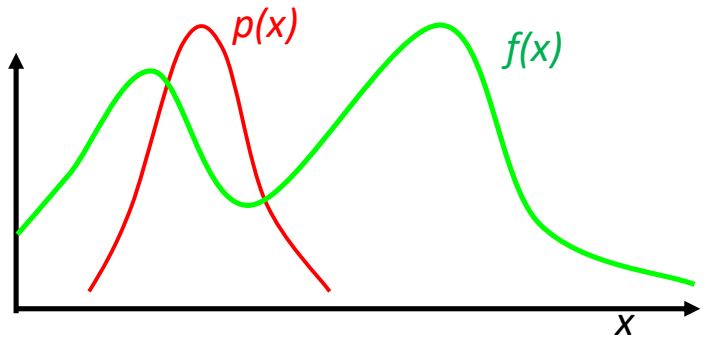




Approximate Expectation

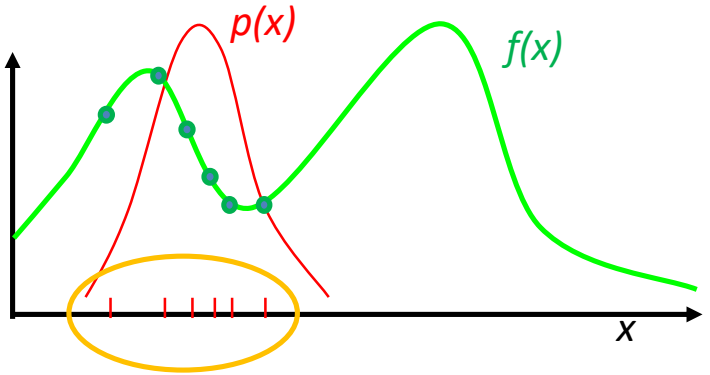
Exact Expectation of $f(x)$ with $x \backsim p(x)$ given by a function

$$E[f] = \int p(x) f(x) dx$$



Approximate Expectation of $f(x)$ with $p(x)$ given by samples $x_n \backsim p(x)$

$$E[f] \approx \frac{1}{N} \sum_{1}^{N} f(x_n)$$



Known samples from *p(x)*

This is called: **MonteCarlo Approximation**

Approximate Expectation

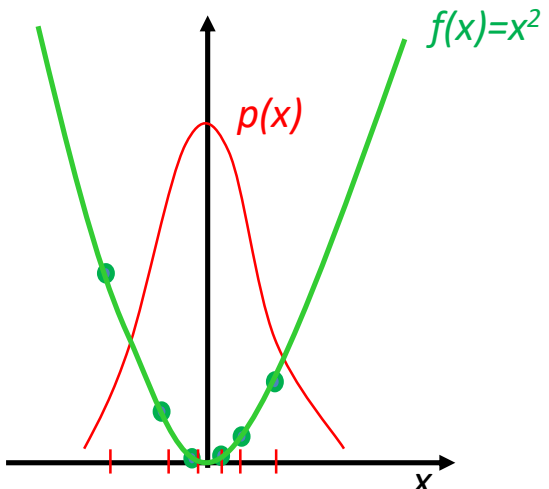
Example:

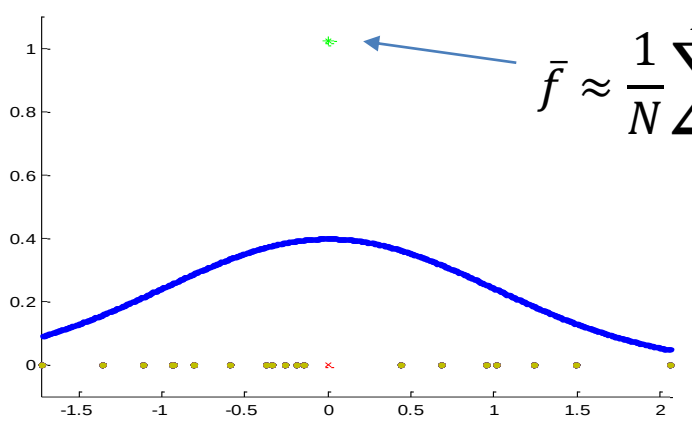
$$f(x) = x^2 \qquad p(x) = N(x; \mu = 0, \sigma^2 = 1)$$

Analytically:

$$\bar{f} = \mathrm{E}[x^2] = \int_{-\infty}^{\infty} N(x; \mu, \sigma^2) x^2 dx = \mu^2 + \sigma^2 = 1$$

With MonteCarlo:

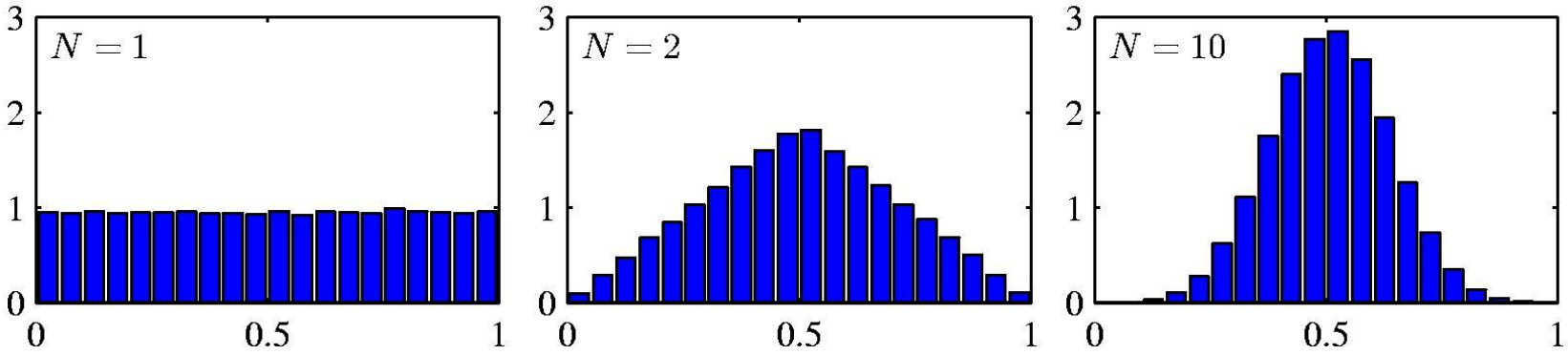




$$\bar{f} \approx \frac{1}{N} \sum_{1}^{N} f(x_n) = 0{,}99$$

Central Limit Theorem

- The distribution of the sum of N i.i.d. random variables becomes increasingly Gaussian as N grows.
- Example: N uniform [0,1] random variables.



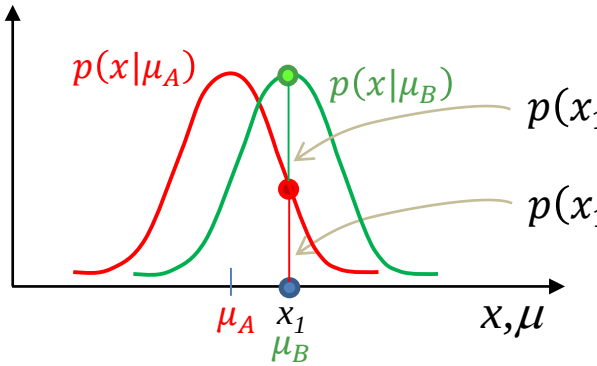
http://blog.vctr.me/posts/central-limit-theorem.html

i.i.d. : independent and identically distributed

Gaussian Parameter Estimation

Given a set of samples **x**={x_1,…, x_n}: which is the most likely gaussian pdf (i.e. μ and σ) from which the samples have been drawn? [*Opposite problem to sampling*]

Example: For just one sample x_1, which is the most likely μ (**given σ**) ?



$$p(x_1|\mu_B) = k\, e^{-\frac{1}{2}\left(\frac{x_1-\mu_B}{\sigma}\right)^2}$$ Probability density for μ_B

$$p(x_1|\mu_A) = k\, e^{-\frac{1}{2}\left(\frac{x_1-\mu_A}{\sigma}\right)^2}$$ Probability density for μ_A

$$p(x_1|\mu_B) > p(x_1|\mu_A)$$

The Gaussian with μ_B gives the highest value → it's the most likely distribution

When we consider $p(x_1|\mu)$ as a function μ (rather than of x) we call it Likelihood function $\mathcal{L}(\mu|x_1) = p(x_1|\mu)$

The key is what is given:

Given μ,σ / function of x : $p(x)=N(x|\mu,\sigma)$, is the probability density of x

Given x,σ/function of μ: $\mathcal{L}(\mu|x,\sigma)=N(x|\mu,\sigma)$ is the likelihood of μ given x, σ

Notice: $\int_{-\infty}^{\infty} p(x|\theta)dx = 1$ but $\int_{-\infty}^{\infty} \mathcal{L}(\theta|x)d\theta \neq 1$ $\theta: parameters\ of\ the\ distribution$

Gaussian Parameter Estimation

Given a set of samples **x**={x_1,…, x_n}: which is the most likely gaussian pdf (i.e. μ and σ) from which the samples have been drawn?

Three options:

1. *Maximum a posteriori* **(MAP):**

$$\arg\max_{\mu,\sigma} p(\mu,\sigma|x) = \arg\max_{\mu,\sigma} p(x|\mu,\sigma)p(\mu,\sigma) = \arg\max_{\mu,\sigma} \mathcal{L}(\mu,\sigma|x)p(\mu,\sigma)$$

2. *Maximum likelihood* **(ML) :** $\arg\max_{\mu,\sigma} p(x|\mu,\sigma) = \arg\max_{\mu,\sigma} \mathcal{L}(\mu,\sigma|x)$

3. ***Bayesian:*** $p(\mu,\sigma|x) = \frac{p(x|\mu,\sigma)p(\mu,\sigma)}{p(x)}$ This estimates a new pdf over μ, σ, not just a single, best μ, σ

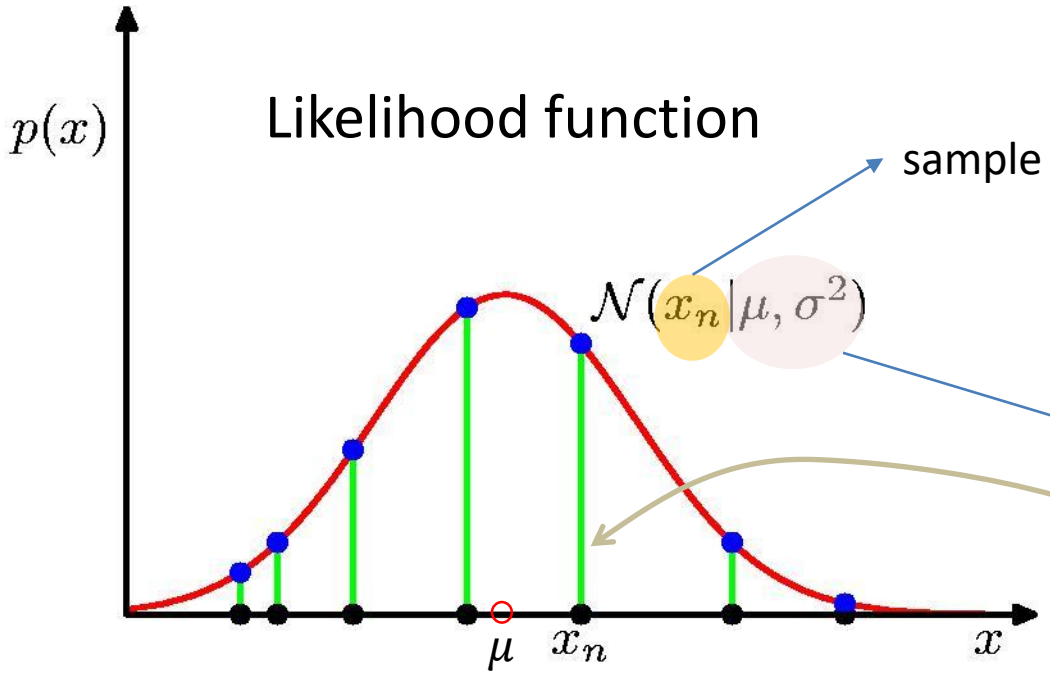
More Next

Gaussian Parameter Estimation

Given a number of samples **x**=$\{x_1,\ldots, x_n\}$ and assuming they are i.i.d.:

$$p(\mathbf{x}|\mu,\sigma^2) = \prod_{n=1}^{N} \mathcal{N}\left(x_n|\mu,\sigma^2\right)$$

Joint probability of the vector **x**



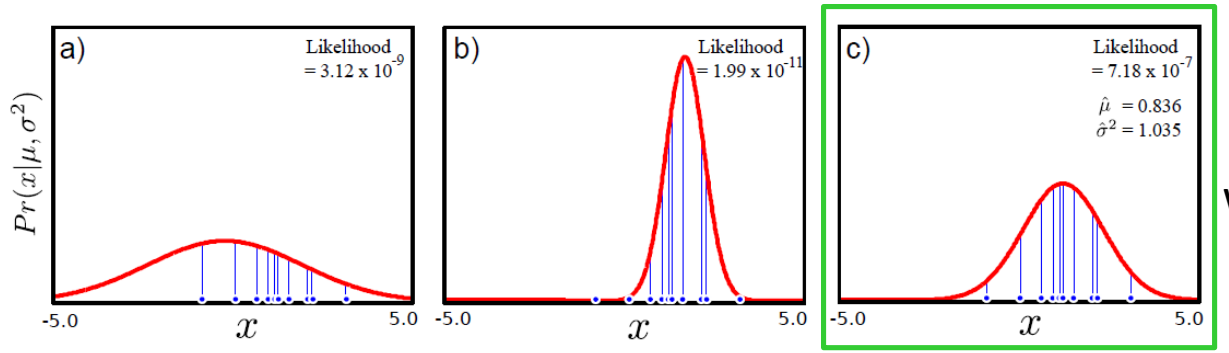
$$k\, e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2} = k\, e^{-\frac{1}{2}MD(x)^2}$$

Gaussian Parameter Estimation

μ_{ML} and σ_{ML} estimation:

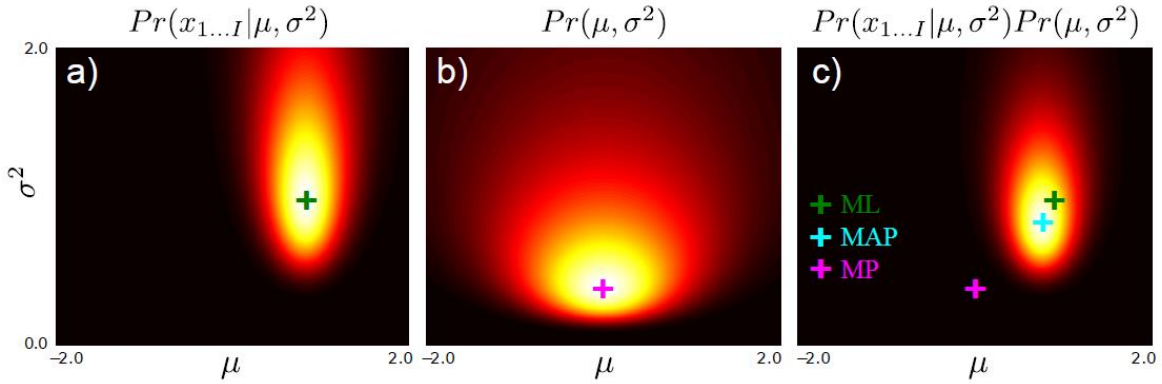
Different parameters μ y σ produce different Likelihood, Prior and MAP

Example of Likelihood: $p(x_{1:10}|\mu)$ for a given data set of 10 data points:



Which is the highest?

Representation in the <μ ,σ>plane :



Likelihood Prior Posterior

(Courtesy: Prince, 2012)

Gaussian Parameter Estimation

Maximum Likelihood (ML) Estimation :

Select the parameters μ_{ML} y σ_{ML} for which the probability of the observed data becomes the highest

$$Arg\ max_{\mu\sigma}\ p(\mathbf{x}|\mu,\sigma^2) = Arg\ max_{\mu\sigma} \prod_{n=1}^{N} \mathcal{N}\left(x_n|\mu,\sigma^2\right)$$

Equivalently: Maximum Log-Likelihood Estimation

$$Arg\ max_{\mu\sigma}\ \ln p\left(\mathbf{x}|\mu,\sigma^2\right) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

SOLUTION:

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n$$ Sample Mean

$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2$$ Sample Variance

Approaching Probabilistic Robotics

Objective: Estimate a vector state x

$$x = \begin{cases} \bullet \text{ robot pose (position and orientation)} \\ \bullet \text{ map} \\ \bullet \text{ robot pose + map (SLAM)} \end{cases}$$

Why probability:

- Inaccurate model of the world (map)
- Uncertainty in the sensor observation
- Uncertainty in the robot motion

Two general approaches:

- Parameter estimation: estimate the best vector x → **Batch** (optimization) **process**
- Bayes Filter: estimate the best probability function over x → **Recursive process**

Probabilistic Robotics as Parameter Estimation

Given an observation z and a likelihood function $p(z|x)$ (not necessary Gaussian)

posterior $\propto$ likelihood × prior

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

Maximum Likelihood (ML) Estimation:

Find x that maximizes the likelihood function $\mathcal{L} \triangleq p(z|x)$

$$x_{ML} = \arg\max_x p(z|x)$$

Example: Estimate the robot position in a corridor given all the measurements $z=\{z_1, \ldots, z_n\}$ at once

Maximum a Posteriori (MAP) Estimation:

Find x that maximizes the posterior probability $p(x|z)$

$$x_{MAP} = \arg\max_x p(z|x)p(x)$$

- If the **prior** $p(x)$ **is non-informative** (very high covariance), then MAP and ML are the same.
- Both ML and MAP compute the **Mode** of a distribution.
- If the distributions are Gaussian, the **Mode = Mean**

Probabilistic Robotics as Bayes Filter

Explicit distinction between the two robot essential operations

- **Sensing (observations z)** → Reduces uncertainty

  Applied with the **BAYES RULE**: Product of distribution

$$p(x_t \mid z_1, \ldots, z_t) = \eta_t \; p(z_t \mid x_t) \; p(x_t \mid z_1, \ldots, z_{n-1})$$

- **Robot motion u** → Increases uncertainty

  Gives us the new pose $\boldsymbol{x_t}$ when executing the motion command $\boldsymbol{u_t}$ and when its current pose is $\boldsymbol{x_{t-1}}$

  Applied with the **TOTAL PROBABILITY law**

$$p(x_t \mid u_{1:t}) = \int p(x_t \mid u_t, x_{t-1}) \; p(x_{t-1} \mid u_{1:t-1}) \, dx_{t-1}$$

pose at time *t*

pose at time *t-1*

motion at *t-1*

Bayes Filter. Observation + Motion

z = observation
u = motion
x = state

$$Belief(x_t) = p(x_t \mid u_1, z_1 \ldots, u_t, z_t) = p(x_t \mid u_{1:t}, z_{1:t})$$

$$z_{1:t} = z_1, \ldots, z_t$$
$$u_{1:t} = u_1, \ldots, u_t$$

Bayes $$= \eta \; p(z_t \mid x_t, \cancel{u_{1:t}, z_{1:t-1}}) \; p(x_t \mid u_{1:t}, z_{1:t-1})$$ ← without z_t

Markov $$= \eta \; p(z_t \mid x_t) \; p(x_t \mid u_{1:t}, z_{1:t-1})$$

$$p(x_t) = \int p(x_t, x_{t-1}) dx_{t-1} = \int p(x_t \mid x_{t-1}) p(x_{t-1}) dx_{t-1}$$

Total prob. $$= \eta \; p(z_t \mid x_t) \int p(x_t \mid u_{1:t}, z_{1:t-1}, x_{t-1}) p(x_{t-1} \mid u_{1:t}, z_{1:t-1}) \, dx_{t-1}$$

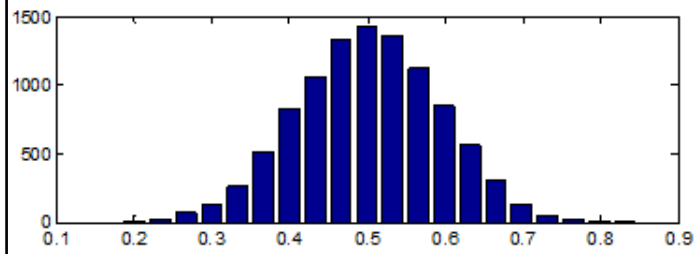
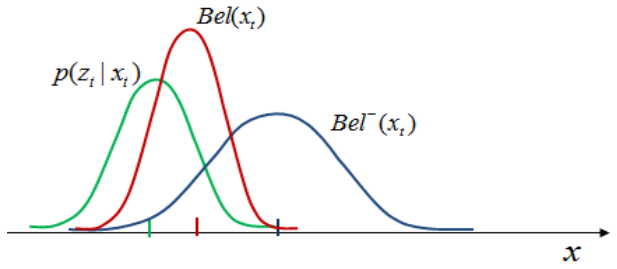
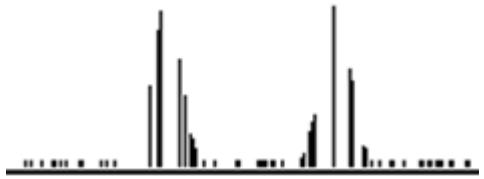
Markov $$= \eta \; p(z_t \mid x_t) \int p(x_t \mid u_t, x_{t-1}) \; p(x_{t-1} \mid u_{1:t}, z_{1:t-1}) \, dx_{t-1}$$

Markov $$= \eta \; p(z_t \mid x_t) \int p(x_t \mid u_t, x_{t-1}) \boxed{p(x_{t-1} \mid u_{1:t-1}, z_{1:t-1})} \, dx_{t-1}$$

We don't apply Markov here because it is not needed. This is the definition of $Belief(x_{t-1})$

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

Bayes Filter can be implemented in a number of ways

	$Bel(x_t)$	Variable x_t	Example 1 D
Discrete (histogram) Filter	Histogram /Grid	Discrete	
Kalman Filter	Gaussian	Continuous	
Particle Filter	Particles	Discrete	

Summary

Three very important rules:

Sum Rule (Marginalization): $P(X) = \sum_Y P(X, Y)$

Product Rule: $P(X, Y) = P(Y \mid X) P(X)$

Law of total probability: $P(X) = \sum_Y P(X \mid Y) P(Y)$

Bayes' Theorem: $P(Y \mid X) = \frac{P(X \mid Y) P(Y)}{P(X)}$

posterior $\propto$ likelihood × prior

Summary

Gaussian Distributions

$$X_1 \sim N(\mu_1, \Sigma_1) = g_1(x)$$ ← Gaussian function 1

$$X_2 \sim N(\mu_2, \Sigma_2) = g_2(x)$$ ← Gaussian function 2

Product (Bayes rule): $g_3(x) = g_1(x)g_2(x) = N(\Sigma_{12}\left(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2\right), \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)^{-1})$

Linear transformation of RVs: $Y = AX + B$ $\quad Y \sim N(A\bar{X} + B, A\Sigma A^t)$

Non- Linear transformation of RVs: $Z = f(X, Y)$

$$Z \sim N(f(\bar{X}, \bar{Y}), \Sigma_Z) \qquad \Sigma_Z = \frac{\partial f}{\partial X}\Sigma_X\left(\frac{\partial f}{\partial X}\right)^T + \frac{\partial f}{\partial Y}\Sigma_Y\left(\frac{\partial f}{\partial Y}\right)^T$$

Summary

Bayes Filter:

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

Posterior — Likelihood — $Bel^{-}(x_t)$ or prior

Three rules are applied:

Bayes' Theorem: $$p(x \mid z) = \frac{p(z \mid x)\, p(x)}{p(z)} = \eta\, p(z \mid x)\, p(x)$$

Law of total probability: $$p(x) = \int p(x \mid y)\, p(y)\, dy$$

Markov assumption: $$p(z_t \mid x_{t-1}, z_{1:t-1}) = p(z_t \mid x_{t-1})$$

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

Summary

1. We will represent the robot's belief by a probability distribution over possible positions and use Bayes Filter to update the belief whenever the robot senses (Bayes rule) or moves (total probability rule)
2. Bayes Filter applies for continuous and discrete RVs
3. Bayes Filter applies for any kind of distribution, not only Normal (Gaussian)
4. When the observation model (likelihood) and motion model are gaussians → the Bayes Filter is known as Kalman Filter (studied next)
5. Not all Probabilistic Robotics is done using Bayes Filter: Batch process (like Least Squares) are also applied