COMP9418: Advanced Topics in Statistical Machine Learning

Gaussian Models

Instructor: Gustavo Batista

University of New South Wales

Introduction

- This lecture discusses Graphical Models with continuous variables
  - We will focus on Gaussian distributions and formalise a Gaussian Bayesian network
  - Our findings can be adapted to other models such as Markov networks

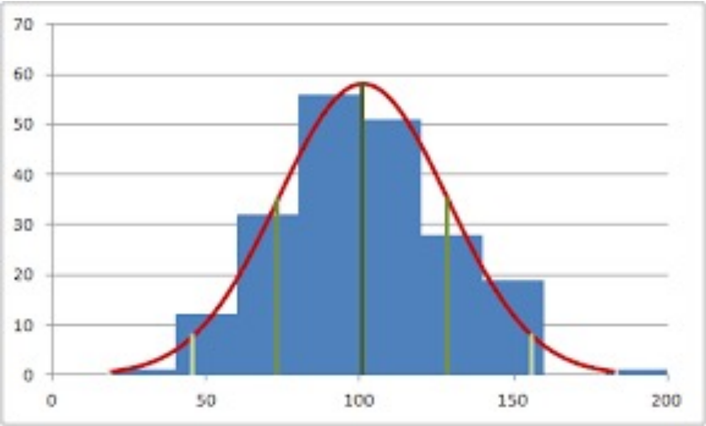
- We will see that our existing knowledge about probabilities applies to continuous variables

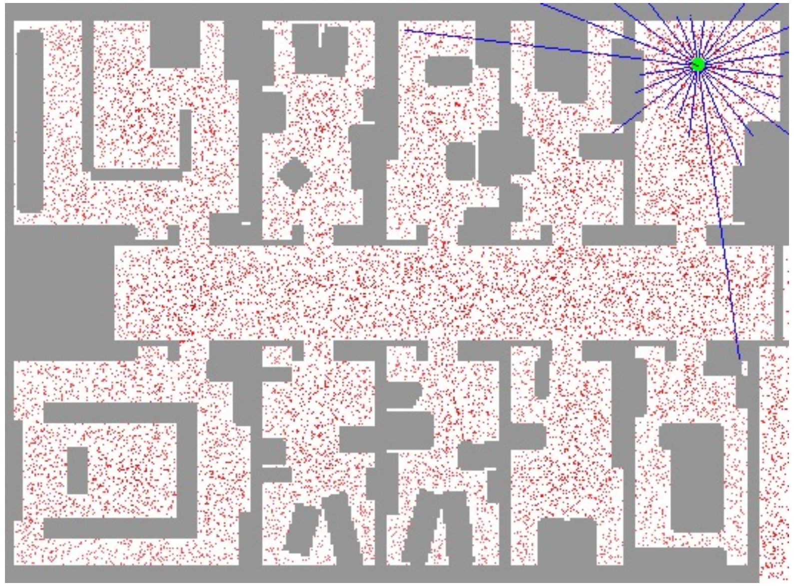
  - Independence, conditional independence
  - Bayes conditioning, product rule, chain rule, case analysis, Bayes rule, etc.

- We will develop a representation for Gaussian Factors
  - Including operations such as join, marginalisation and reduction (observation of evidence)
  - We will use these operations to illustrate how Kalman Filters works

Introduction

- Let’s now see how we can incorporate continuous variables in our models
  - Some variables are best modelled in the continuous space, such as temperature, humidity, position and velocity.
  - We cannot use tables anymore, unless we discretise the variables
- Discretisation is a common approach
  - We can approximate a variables distribution by its histogram
  - But it is hardly the answer for all models
- Imagine the problem of robot navigation
  - A large environment and a resolution of 15 x 15 cm would lead to millions of values
  - Such large CPTs would be too expensive to make inference
  - Besides, we lose the notion of distance between values





Introduction

- First, everything we know about probabilities holds for continuous distributions
  - Bayes conditioning and product rule
  - Chain rule
  - Case analysis and marginalisation
  - Bayes rule

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

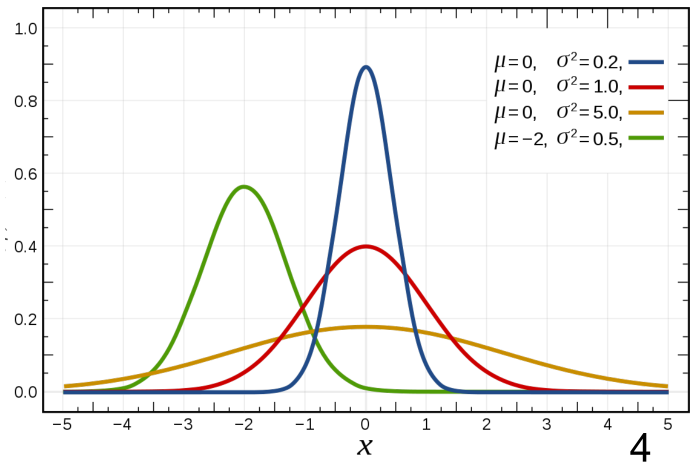
$$P(A,B) = P(A|B)P(B)$$

$$P(A_1, \dots, A_n) = \prod_i P(A_i|A_{i-1}, \dots, A_1)$$

$$P(A) = \int_B P(A,B)dB$$

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

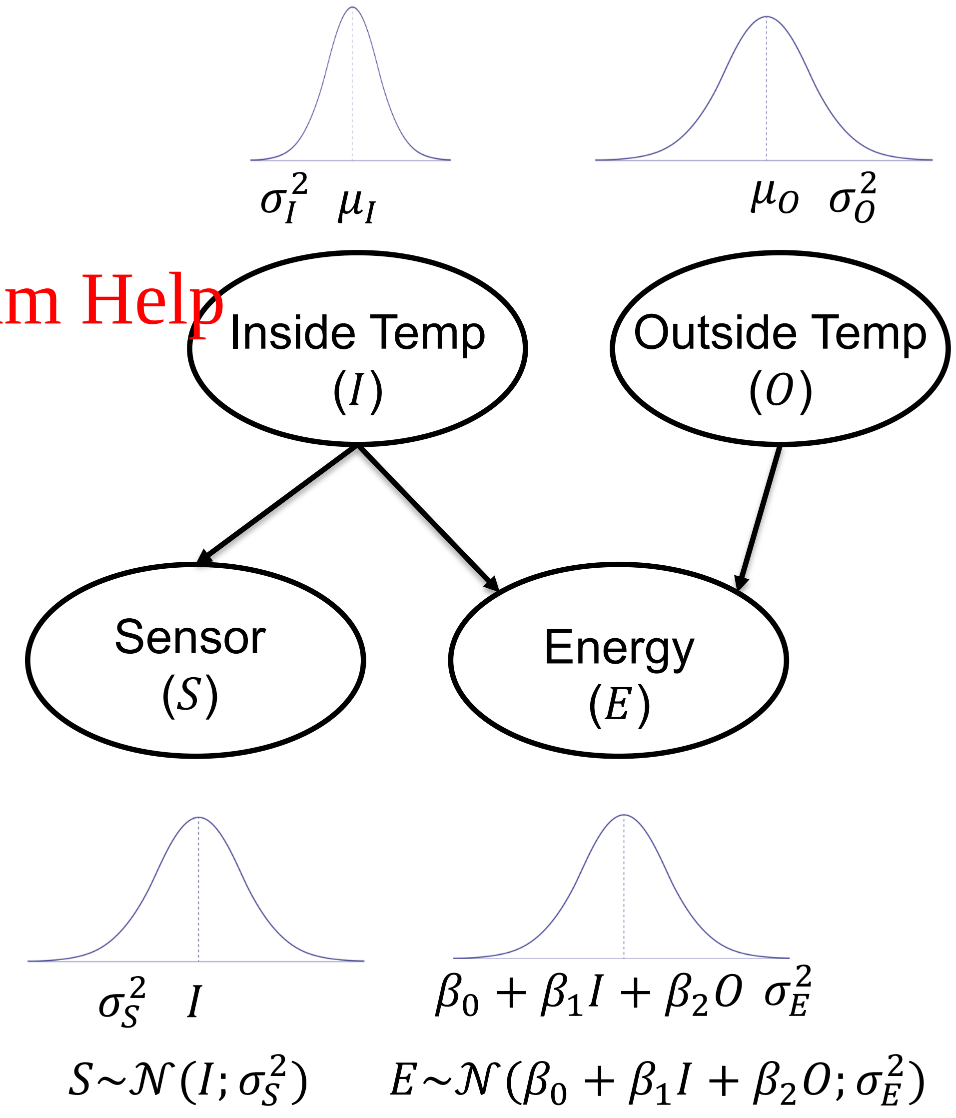
- However, our operations over tables will not work for continuous variables
  - We need to represent the distribution using a *probability density function* (PDF)
  - A common PDF for continuous variables is the Gaussian distribution



Gaussian Bayesian Networks

- Gaussian Bayesian networks
  - All variables are continuous and modelled by Gaussian densities
  - Gaussians are often good approximation for many real-world distributions

- Modelling decisions
  - Root nodes use univariate distributions
  - We need to represent the CPD $P(X|\boldsymbol{U})$, where $\boldsymbol{U}$ are the parents of X
  - A common solution is a *linear Gaussian model*



Linear Gaussian Model

- Let X be a continuous variable with continuous parents $U_1, \dots, U_k$
  - X has a linear Gaussian model with parameters $\beta_0, \dots, \beta_k$ and σ^2 iff
  - Similarly, in vector notation
  - Yet, we can understand X as a linear function of $U_1, \dots, U_k$ with a Gaussian noise with mean 0 and variance σ^2

$$P(X|u_1, \dots, u_k) = \mathcal{N}(\beta_0 + \beta_1 u_1 + \cdots + \beta_k u_k; \sigma^2)$$

$$P(X|\boldsymbol{u}) = \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{u}; \sigma^2)$$

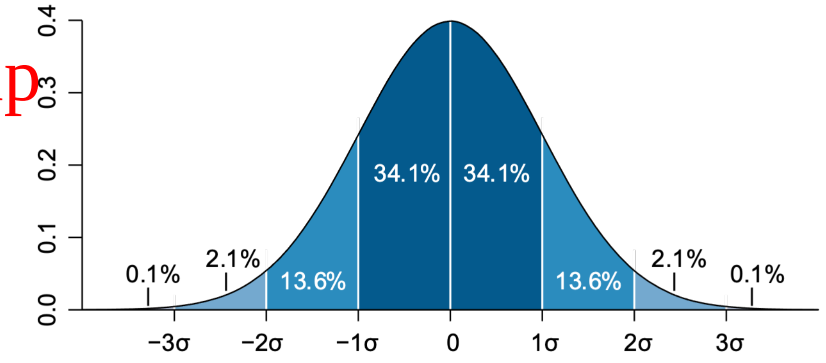
$$X = \beta_0 + \beta_1 u_1 + \cdots + \beta_k u_k + \epsilon$$

- The linear model assumes the variance does not depend on the parents $\boldsymbol{U}$
  - We can easily extend it to have mean and variance of X depend on parents
  - However, linear Gaussian model is a useful approximation in many practical problems.
  - It also provides an alternative representation for multivariable Gaussian distributions

Gaussian Distribution: 1 Dimension

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

- Univariate Gaussian distribution has two parameters
  - Mean μ and
  - Variance σ^2 or standard deviation σ



- Learning is estimating the parameter μ and σ from data
  - $\mu = \mathbb{E}[X]$
  - $\sigma = \sqrt{\mathbb{E}[(X-\mu)^2]}$

- Once, we have learned the parameters, we can sample from the distribution
  - $X \sim \mathcal{N}(\mu; \sigma^2)$

Gaussian Distribution: 2 Dimensions

- Let's suppose we have two independent variables X_1 and X_2

$$p(x_1, x_2) = p(x_1)p(x_2)$$

$$p(x_1, x_2) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}$$

$$= \frac{1}{\sigma_1\sigma_2 2\pi} e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}$$

$$= \frac{1}{\sigma_1\sigma_2 2\pi} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

$$= \frac{1}{\sigma_1\sigma_2 2\pi} e^{-\frac{1}{2}\left((\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)}$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

$$\Sigma^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix}$$

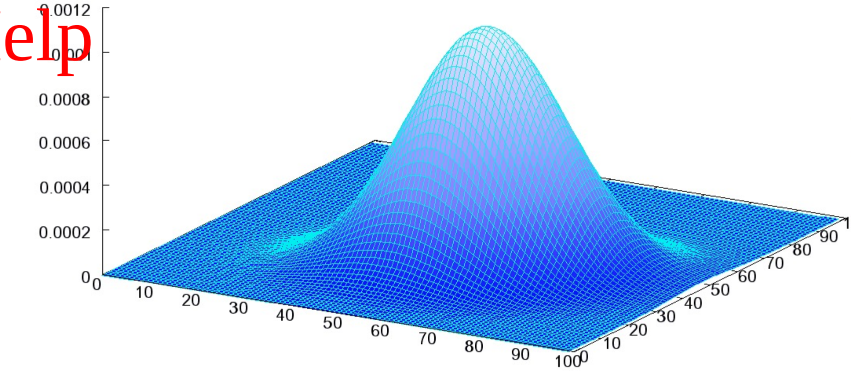
$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

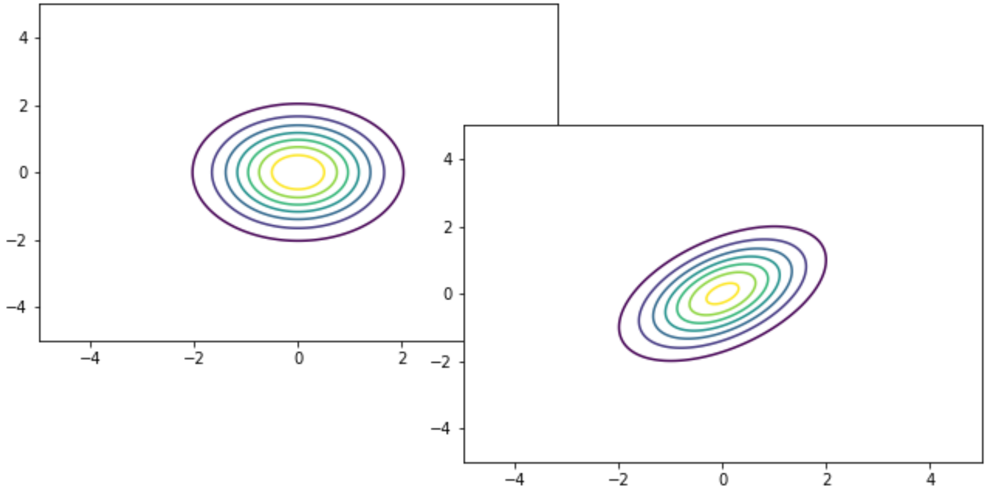
Gaussian Distribution: n Dimensions

- Multivariate Gaussian distribution is characterised by
  - n-dimensional *mean vector* $\boldsymbol{\mu}$ and
  - Symmetric $n \times n$ covariance matrix Σ
  - Quadratic number of parameters

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$



- Covariance matrix for 2 dimensions
  - $cov(X_1, X_2) = \mathbb{E}[(x_1 - \mu_1)(x_2 - \mu_2)]$
  - $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{pmatrix} \sigma_1^2 & cov(X_1, X_2) \\ cov(X_2, X_1) & \sigma_2^2 \end{pmatrix} \right)$
  - Covariance matrix is symmetric since $cov(X_1, X_2) = cov(X_2, X_1)$



Gaussian Distribution: Example

- Consider a joint distribution $p(X_1, X_2, X_3)$ over three variables
  - $\mu = \begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix}$
  - $\Sigma = \begin{pmatrix} 4 & 2 & -2 \\ 2 & 5 & -5 \\ -2 & -5 & 8 \end{pmatrix}$

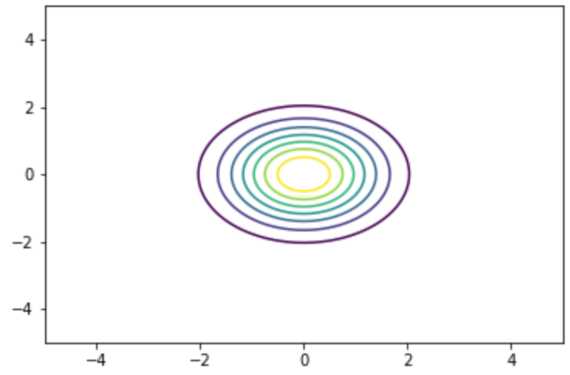
- We can observe that
  - X_1 is positively correlated with X_2
  - X_1 is negatively correlated with X_3
  - X_2 is negatively correlated with X_3

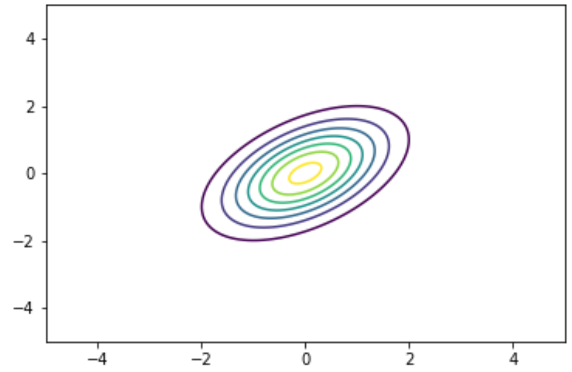
Gaussian Distribution: Independencies

- We can identify independence assumptions directly from the Gaussian distribution parameters
  - X_i and X_j are independent iff and only if $\Sigma_{i,j} = 0$

- Let $J = \Sigma^{-1}$ be the *information matrix*
  - $X_i \perp X_j \mid \boldsymbol{X} - \{X_i, X_j\}$ iff $J_{i,j} = 0$

- Example
  - $\Sigma = \begin{pmatrix} 4 & 2 & -2 \\ 2 & 5 & -5 \\ -2 & -5 & 8 \end{pmatrix} \qquad J = \Sigma^{-1} = \begin{pmatrix} .3125 & -.125 & 0 \\ -.125 & .5833 & .3333 \\ 0 & .3333 & .3333 \end{pmatrix}$
  - $X_1 \perp X_3 \mid X_2$

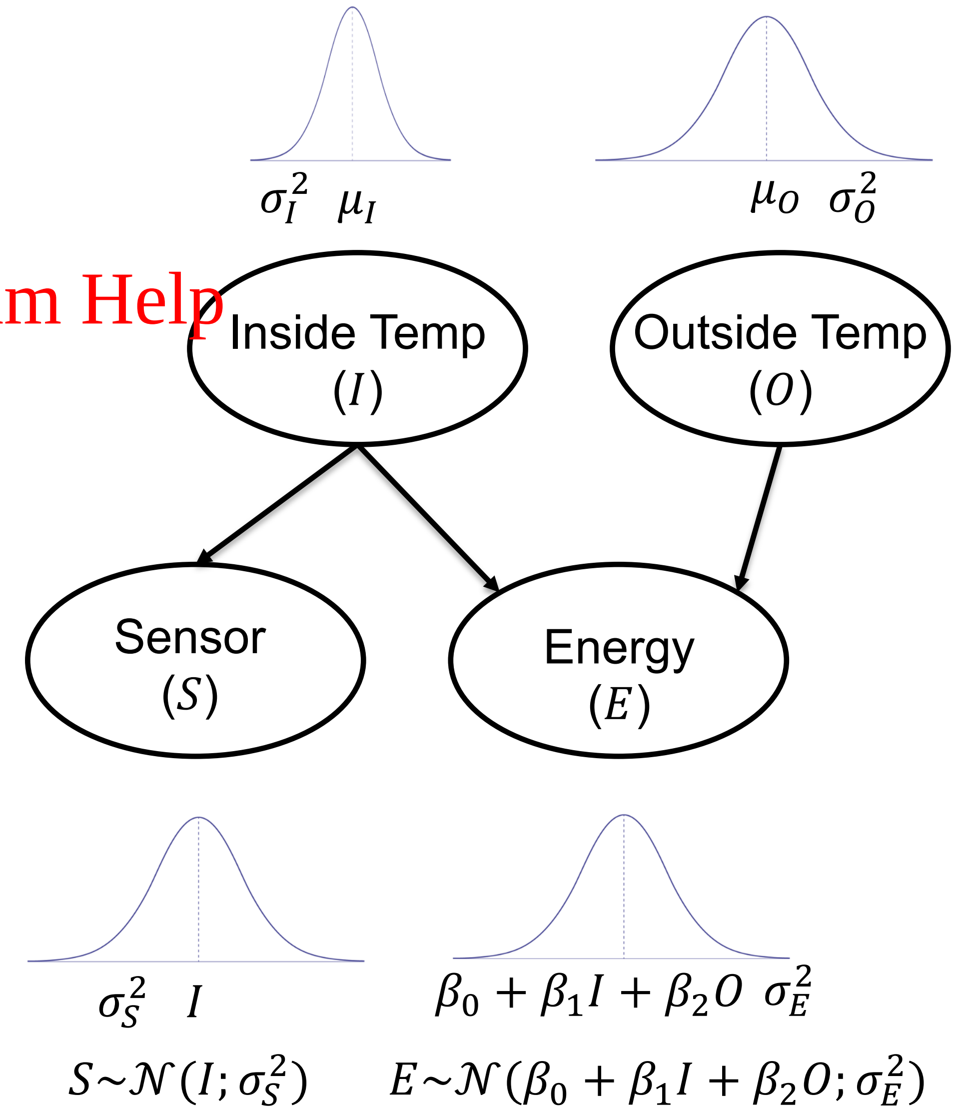




Gaussian Bayesian Networks

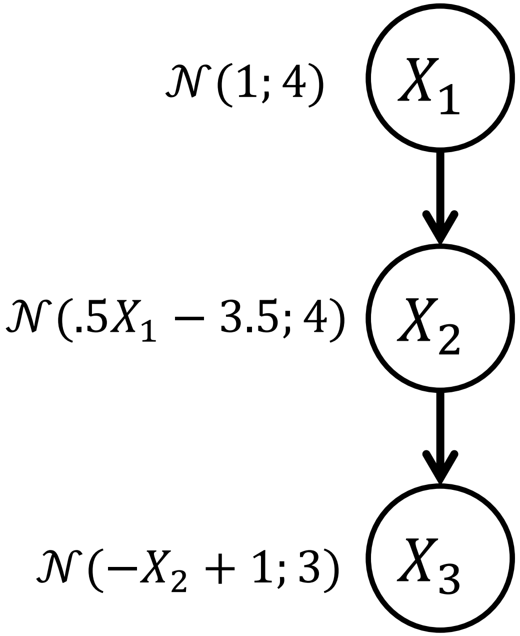
- A Gaussian Bayesian network has
  - All variables are continuous
  - All CPDs are linear Gaussian models

- Gaussian Bayesian networks are simple to understand
  - If we compare to multivariate Gaussian distributions
  - Yet, we can transform one representation into another



GBN and Multivariate Gaussian 1

- A linear Gaussian network defines a joint multivariate Gaussian distribution
  - Y is a linear Gaussian with parents $X_1, \dots, X_k$
  - $P(Y|\boldsymbol{x}) = \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{x}; \sigma^2)$
  - $X_1, \dots, X_k$ are jointly Gaussian with $\mathcal{N}(\boldsymbol{\mu}; \Sigma)$



- Then, Y distribution is normal $p(Y) = \mathcal{N}(\mu_Y; \sigma_Y^2)$, where
  - $\mu_Y = \beta_0 + \boldsymbol{\beta}^T \boldsymbol{\mu}$
  - $\sigma_Y^2 = \sigma^2 + \boldsymbol{\beta}^T \Sigma \boldsymbol{\beta}$

- The joint distribution over $\{\boldsymbol{X}, Y\}$ is normal with
  - $Cov[X_i; Y] = \Sigma_{j=1}^{k} \beta_j \Sigma_{i,j}$

GBN and Multivariate Gaussian 2

- A joint multivariate Gaussian distribution defines a linear Gaussian network
  - Given a set of variables $\{\boldsymbol{X}, Y\}$ in the form of a joint normal distribution
  - $p(Y|\boldsymbol{X}) = \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{X}; \sigma^2)$

- Then,
  - $\beta_0 = \mu_Y - \Sigma_{Y\boldsymbol{X}} \Sigma_{\boldsymbol{XX}}^{-1} \boldsymbol{\mu}_{\boldsymbol{X}}$
  - $\boldsymbol{\beta} = \Sigma_{\boldsymbol{XX}}^{-1} \Sigma_{\boldsymbol{X}Y}$
  - $\sigma^2 = \Sigma_{YY} - \Sigma_{Y\boldsymbol{X}} \Sigma_{\boldsymbol{XX}}^{-1} \Sigma_{\boldsymbol{X}Y}$

$$\Sigma = \begin{pmatrix} 4 & 2 & -2 \\ 2 & 5 & -5 \\ -2 & -5 & 8 \end{pmatrix}$$

$$\mu = \begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix}$$

Gaussian Bayesian Networks

- Given our knowledge about inference, we need the following operations
  - Multiply factors
  - Marginalise out variables (using integration)
- Multiplication
  - We do not have a universal representation, as we have for discrete distributions
  - Multiplication of factors of different families is difficult
  - Multiplication of factors in the same family may lead to results in a different family
- Marginalisation
  - Not all functions are integrable
  - If they are, not all have a closed-form integral

Canonical Form

- We need to adopt a representation that allow us perform inference operations in closed form
  - A simple option is the *canonical form*
  - Factor product, reduction and marginalisation in closed form

- We can define a data structure that stores factors in the canonical form
  - The canonical form can represent multidimensional Gaussian distributions and linear Gaussian CPDs
  - Adapt inference algorithms, such as VE, to operate over this new factor

Canonical Form

- The canonical form $\mathcal{C}(\boldsymbol{X}; K, \boldsymbol{h}, g)$ is defined as

$$\mathcal{C}(\boldsymbol{X}; K, \boldsymbol{h}, g) = \exp\left(-\frac{1}{2}\boldsymbol{X}^T K \boldsymbol{X} + \boldsymbol{h}^T \boldsymbol{X} + g\right)$$

- Thus, $\mathcal{N}(\boldsymbol{\mu}; \Sigma) = \mathcal{C}(K, \boldsymbol{h}, g)$, where
  - $K = \Sigma^{-1}$
  - $\boldsymbol{h} = \Sigma^{-1}\boldsymbol{\mu}$
  - $g = -\frac{1}{2}\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu} - \log\left((2\pi)^{n/2}|\Sigma|^{1/2}\right)$

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

Canonical Form: Join

- The product of two canonical form factors over scope $\boldsymbol{X}$ is
  - If the factors have different scopes, we extend the scopes to make them match
  - The extension of scope is simply adding zero entries to K and $\boldsymbol{h}$

$$\mathcal{C}(K_1, \boldsymbol{h}_1, g_1) \cdot \mathcal{C}(K_2, \boldsymbol{h}_2, g_2) = \mathcal{C}(\mathrm{K}_1 + \mathrm{K}_2, \mathbf{h}_1 + \mathbf{h}_2, \mathrm{g}_1 + \mathrm{g}_2)$$

- Let us compute $\phi_1(X,Y) \cdot \phi_2(Y,Z)$
  - $\phi_1(A,B) = \mathcal{C}\left(A, B; \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, -3\right)$
  - $\phi_2(B,C) = \mathcal{C}\left(B, C; \begin{bmatrix} 3 & -2 \\ -2 & 4 \end{bmatrix}, \begin{pmatrix} 5 \\ -1 \end{pmatrix}, 1\right)$
  - $\phi_1(A,B) \cdot \phi_2(B,C) = \mathcal{C}\left(A, B, C; \begin{bmatrix} 1 & -1 & 0 \\ -1 & 4 & -2 \\ 0 & -2 & 4 \end{bmatrix}, \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix}, -2\right)$

Canonical Form: Marginalisation

- The marginalisation of $\boldsymbol{Y}$ for a canonical form $\mathcal{C}(\boldsymbol{X},\boldsymbol{Y}; K_1, \boldsymbol{h}_1, g_1)$ over scope $\{\boldsymbol{X},\boldsymbol{Y}\}$ is
  - $K = \begin{bmatrix} K_{\boldsymbol{XX}} & K_{\boldsymbol{XY}} \\ K_{\boldsymbol{YX}} & K_{\boldsymbol{YY}} \end{bmatrix}$
  - $\boldsymbol{h} = \begin{pmatrix} \boldsymbol{h}_{\boldsymbol{X}} \\ \boldsymbol{h}_{\boldsymbol{Y}} \end{pmatrix}$

$$K' = K_{\boldsymbol{XX}} - K_{\boldsymbol{XY}} K_{\boldsymbol{YY}}^{-1} K_{\boldsymbol{YX}}$$
$$\boldsymbol{h}' = \boldsymbol{h}_{\boldsymbol{X}} - K_{\boldsymbol{XY}} K_{\boldsymbol{YY}}^{-1} \boldsymbol{h}_{\boldsymbol{Y}}$$
$$g' = g + \frac{1}{2}\left(\log\left|2\pi K_{\boldsymbol{YY}}^{-1}\right| + \boldsymbol{h}_{\boldsymbol{Y}}^T K_{\boldsymbol{YY}}^{-1} \boldsymbol{h}_{\boldsymbol{Y}}\right)$$

- Let us compute $\int \mathcal{C}(A,B,C; K, \boldsymbol{h}, g)dC$
  - $\mathcal{C}\left(A,B,C; \begin{bmatrix} 1 & -1 & 0 \\ -1 & 4 & -2 \\ 0 & -2 & 4 \end{bmatrix}, \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix}, -2\right)$

Canonical Form: Reduction

- The reduction of a canonical form $\mathcal{C}(\boldsymbol{X}, \boldsymbol{Y}; K_1, \boldsymbol{h}_1, g_1)$ by setting evidence $\boldsymbol{Y} = \mathbf{y}$ is
  - $K = \begin{bmatrix} K_{\boldsymbol{XX}} & K_{\boldsymbol{XY}} \\ K_{\boldsymbol{YX}} & K_{\boldsymbol{YY}} \end{bmatrix}$
  - $h = \begin{pmatrix} h_{\boldsymbol{X}} \\ h_{\boldsymbol{Y}} \end{pmatrix}$

$$K' = K_{\boldsymbol{XX}}$$
$$\boldsymbol{h}' = \boldsymbol{h}_{\boldsymbol{X}} - K_{\boldsymbol{XY}}\boldsymbol{y}$$
$$g' = g + \boldsymbol{h}_{\boldsymbol{Y}}^T\boldsymbol{y} - \frac{1}{2}\boldsymbol{y}^T K_{\boldsymbol{YY}}\boldsymbol{y}$$

- Let us set $C = 2$ for $\mathcal{C}(A, B, C; K, \boldsymbol{h}, g)$
  - $\mathcal{C}\left(A, B, C; \begin{bmatrix} 1 & -1 & 0 \\ -1 & 4 & -2 \\ 0 & -2 & 4 \end{bmatrix}, \begin{pmatrix} 1 \\ 4 \\ -1 \end{pmatrix}, -2\right)$

Canonical Form: Linear Model

- The linear Gaussian model
  - $Y \sim \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{X}; \sigma^2)$

$$K_{\boldsymbol{Y}|\boldsymbol{X}} = \begin{bmatrix} {}^{1}/_{\sigma^2} & -{}^{1}/_{\sigma^2}\,\boldsymbol{\beta}^T \\ -{}^{1}/_{\sigma^2}\,\boldsymbol{\beta} & {}^{1}/_{\sigma^2}\,\boldsymbol{\beta}\boldsymbol{\beta}^T \end{bmatrix}$$

$$\boldsymbol{h}_{\boldsymbol{Y}|\boldsymbol{X}} = \begin{pmatrix} {}^{1}/_{\sigma^2}\,\beta_0 \\ -{}^{1}/_{\sigma^2}\,\beta_0\boldsymbol{\beta} \end{pmatrix}$$

$$g_{\boldsymbol{Y}|\boldsymbol{X}} = -\frac{1}{2}\left(\frac{\beta_0^2}{\sigma^2}\right) - \frac{1}{2}\log(2\pi\sigma^2)$$

Variable Elimination and Gaussian Models

Input: Bayesian network N, query variables $\boldsymbol{Q}$, variable ordering π, evidence $\boldsymbol{e}$

Output: joint marginal $P(\boldsymbol{Q}, \boldsymbol{e})$

1: $\boldsymbol{S} \leftarrow \{f^{\boldsymbol{e}}: f$ is a CPDs of network $N\}$

2: **for** $i = 1$ to length of order π **do**

3: $\quad f \leftarrow \prod_k f_k$ where f_k belongs to $\boldsymbol{S}$ and mentions variable $\pi(i)$

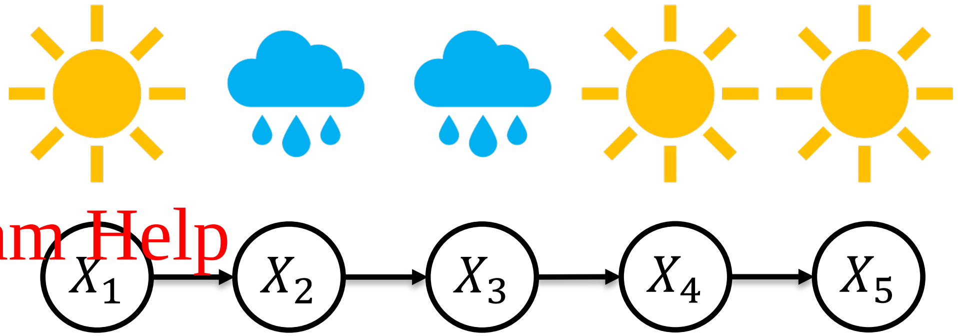
4: $\quad f_i \leftarrow \sum_{\pi(i)} f$

5: $\quad$ replace all factors f_k in $\boldsymbol{S}$ by factor f_i

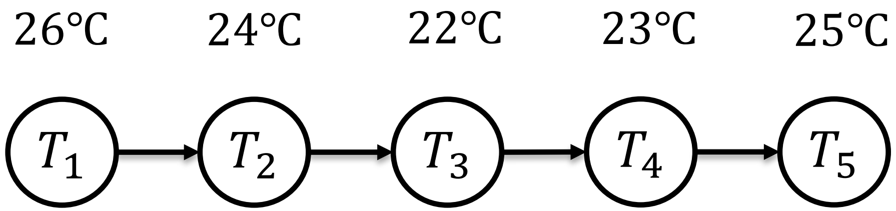
6: **return** $\prod_{f \in \boldsymbol{S}} f$

Kalman Filter

- A *Kalman Filter* is a Hidden Markov Model with continuous variables
  - Root nodes are modelled with Gaussian distributions
  - Internal nodes are linear Gaussian models
  - Thus, Kalman Filter is a Gaussian Bayesian network

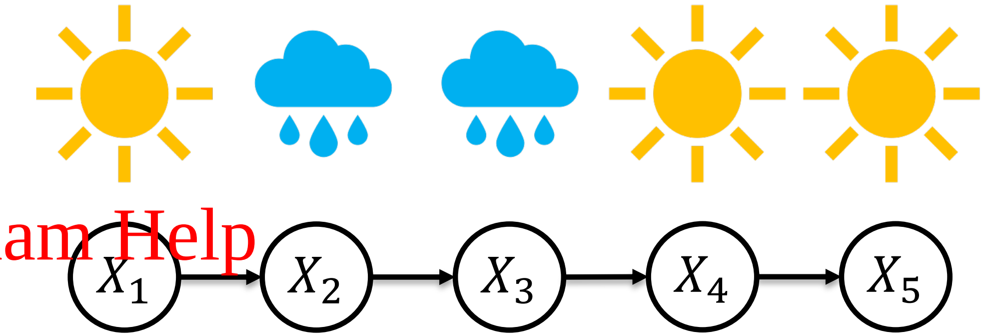


- Let's start with a Markov chain
  - Instead of tracking discrete states such as sun and rain
  - We will track a continuous variable such as temperature

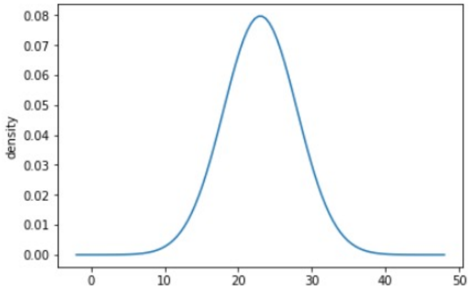


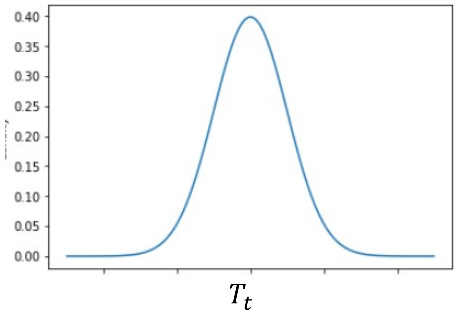
Kalman Filter

- A *Kalman Filter* is a Hidden Markov Model with continuous variables
  - Root nodes are modelled with Gaussian distributions
  - Internal nodes are linear Gaussian models



$$T_t \longrightarrow T_{t+1}$$





$\mu_{T_1} = 23°\text{C}$
$\sigma_{T_1} = 5°\text{C}$

$\mu_{T_{t+1}} = T_t$
$\sigma_{T_{t+1}} = 1°\text{C}$

Kalman Filter

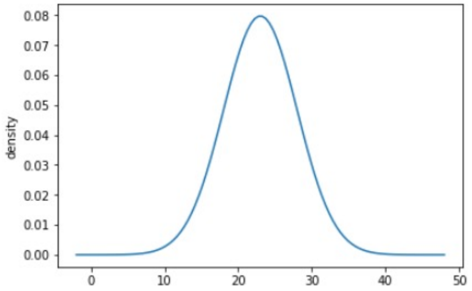
- A *Kalman Filter* is a Hidden Markov Model with continuous variables
  - Root nodes are modelled with Gaussian distributions
  - Internal nodes are linear Gaussian models

μ: 23°C 23°C 23°C 23°C 23°C

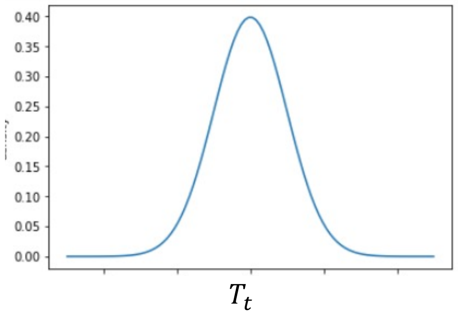
$$\begin{tikzcd} T_1 \arrow[r] & T_2 \arrow[r] & T_3 \arrow[r] & T_4 \arrow[r] & T_5 \end{tikzcd}$$

σ: 5°C 5.1°C 5.2°C 5.3°C 5.4°C

$$\begin{tikzcd} T_t \arrow[r] & T_{t+1} \end{tikzcd}$$



$\mu_{T_1} = 23°\mathrm{C}$
$\sigma_{T_1} = 5°\mathrm{C}$

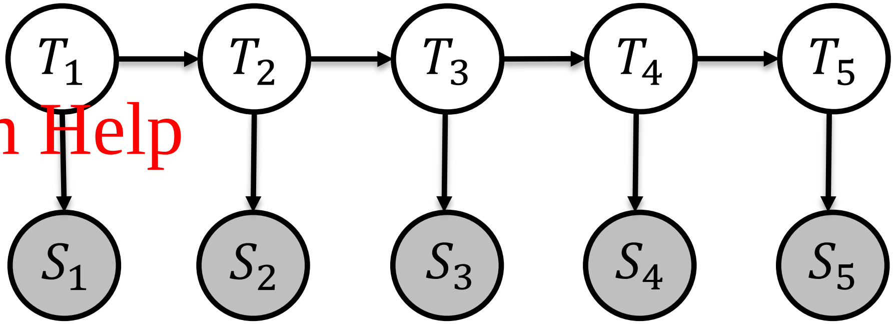


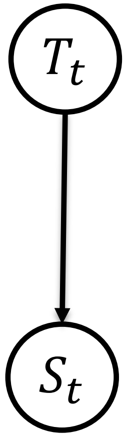
$\mu_{T_{t+1}} = T_t$
$\sigma_{T_{t+1}} = 1°\mathrm{C}$

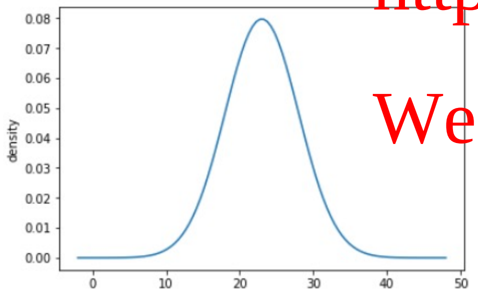
$$p(T_{t+1}) = \int_{T_t} p(T_{t+1}|T_t)p(T_t)dT_t$$

Kalman Filter

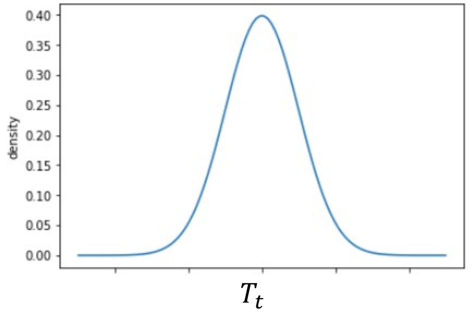
- A *Kalman Filter* is a Hidden Markov Model with continuous variables
  - Root nodes are modelled with Gaussian distributions
  - Internal nodes are linear Gaussian models







$\mu_{T_1} = 23°C$
$\sigma_{T_1} = 5°C$



$\mu_{S_t} = T_t$
$\sigma_{S_t} = .5°C$

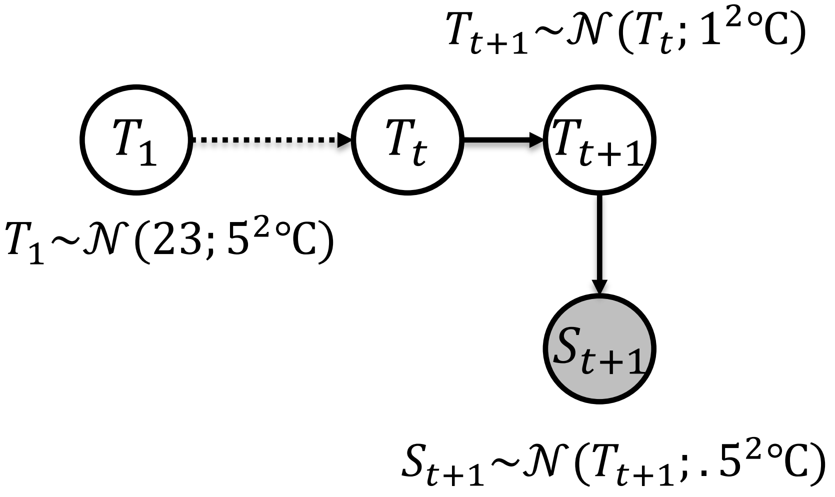
$$p(T_{t+1}) = \int_{T_t} p(T_{t+1}|T_t)p(T_t)dT_t$$

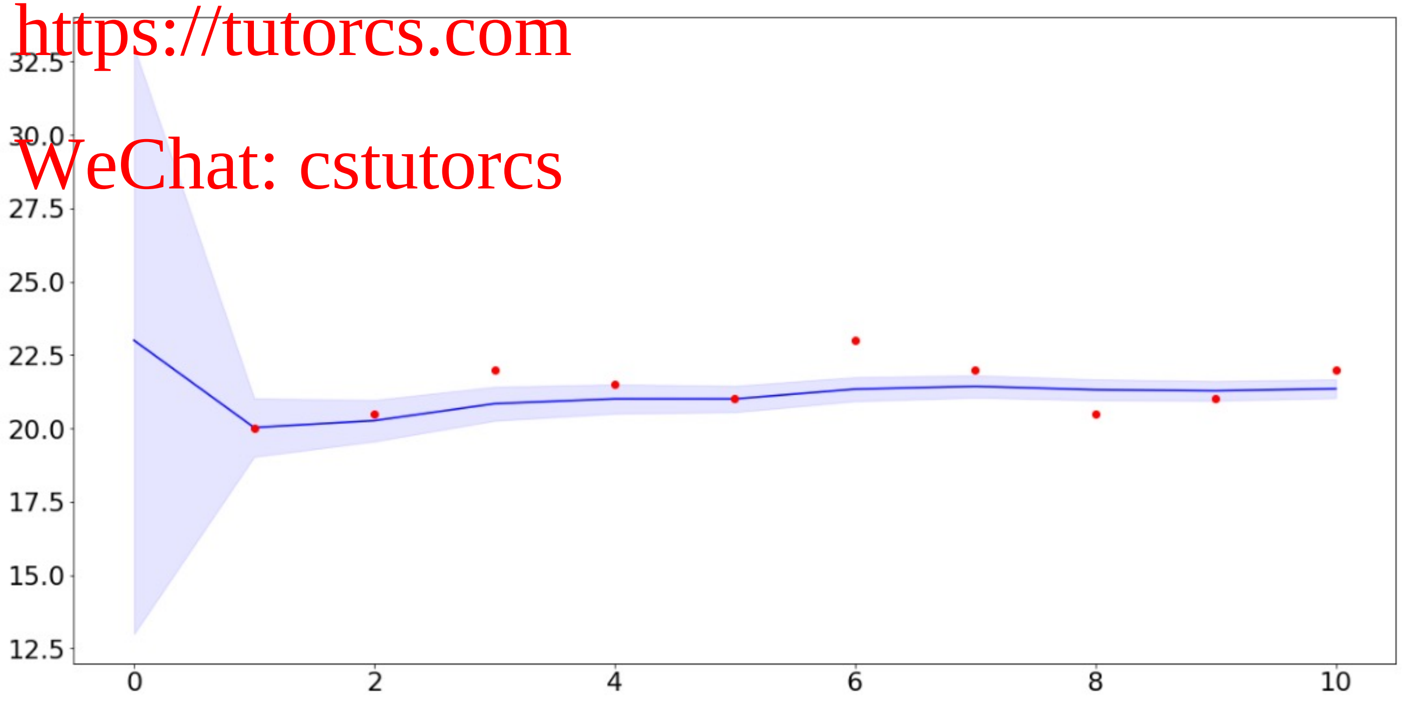
$$p(T_{t+1}|s_{t+1}) \propto p(s_{t+1}|T_{t+1})p(T_{t+1})$$

We must renormalise the results

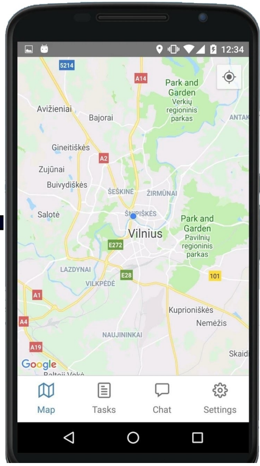
Kalman Filter

- Let's simulate this algorithm with the following evidence:
  - $e = [20, 20.5, 22, 21.5, 21, 23, 22, 20.5, 21, 22]$



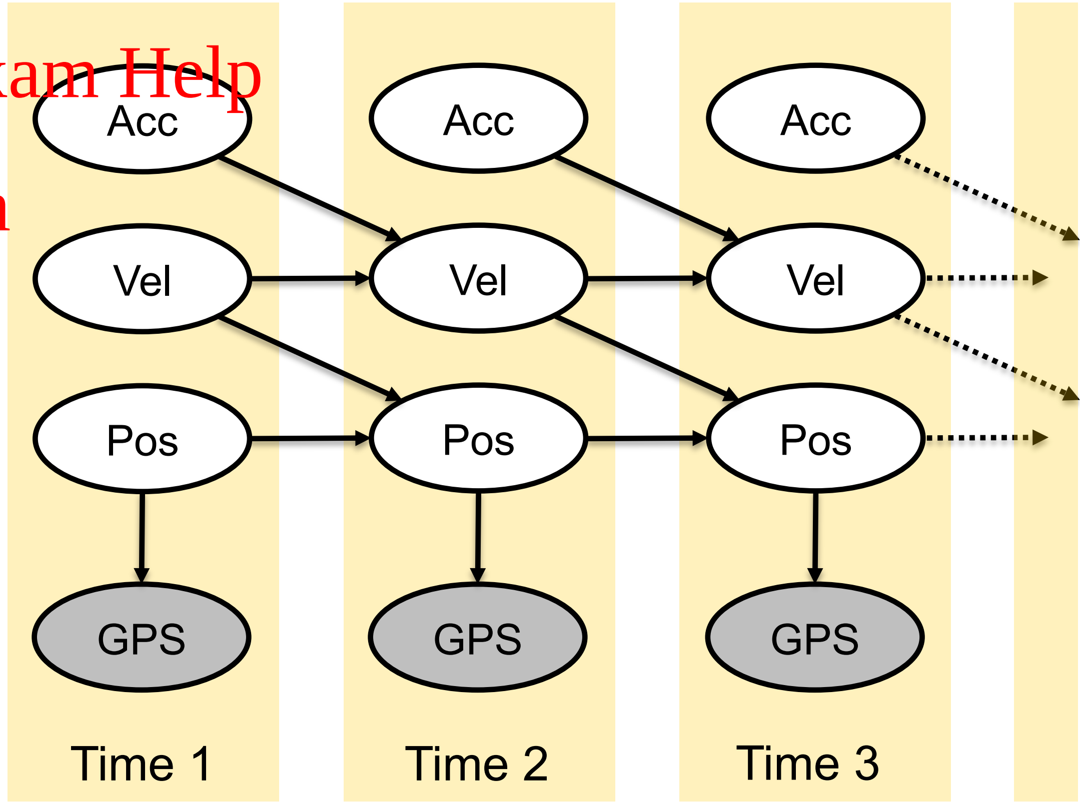


Kalman Filter: GPS



- A Kalman Filter is the core algorithm of GPS systems
  - Example of *data fusion* algorithm
  - We can have additional observations such as the phone accelerometer

- Kalman filters are often applied to guidance and navigations systems
  - Initially applied to trajectory estimation for the Apollo program
  - Currently used in missiles and spacecraft navigation systems, including the International Space Station



Conclusion

- This lecture discussed Graphical Models with continuous variables
  - We used the normal distribution for root variables and the linear Gaussian model for CPDs
  - The canonical representation allows efficient implementation of operations
  - Generally, the use of different continuous distributions is very challenging
- There are several possible extensions
  - Hybrid networks mix continuous and Gaussian variables
  - Nonlinear models such as Extended and Unscented Kalman filter can provide better models when the linear Gaussian model is not appropriated
- The material of this lecture is spread in multiple chapters of Koller & Friedman
  - Continuous variables 5.5, pg. 185-190
  - Gaussian Networks 7.1 & 7.2, pgs. 247-253
  - Variable Elimination in Gaussian Networks, 14.1 & 14.2, pgs. 605-614