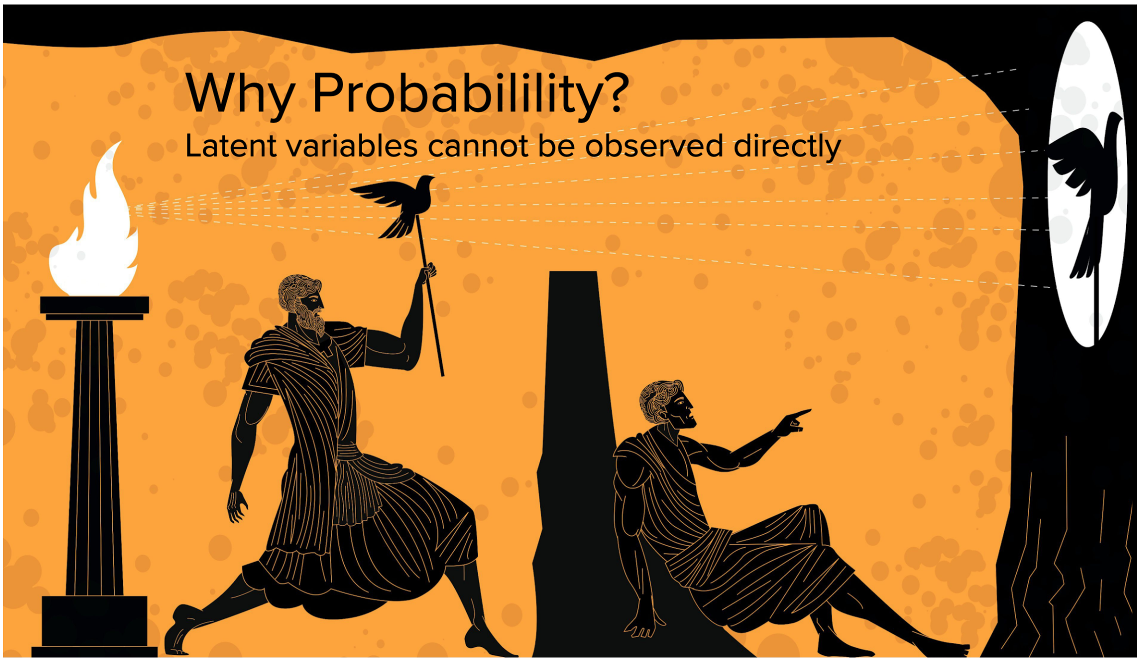
# LGBIO2060 - Pratical sessions

Guillaume Deside

Q7 - 2021



# Contents

# 1 Introduction

#### Author's note

This synthesis is based on the exercise sessions of the LGBIO2060 course given by Clémence Vandamme and Simon Vandergooten. I've tried to be as exact and clear as possible, and to provide as much information as feasible. This synthesis is intended to be as thorough as possible, but it may still be incomplete. If you see any glaring errors, improper visuals, or severe spelling problems, please notify me through Messenger or at my email address: **guillaume.deside@student.uclouvain.be**.

# 2 TP1 - Probability Distributions and Statistical Inference

Equation for a **Gaussian probability** density function :

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

**Some important Definitions:**

- $P(A) \in [0,1]$ is the probability that event A is true.
- $P(\neg A) = 1 - P(A)$ is the **complementary probability** of the event A.
- The text in the entries may be of any length.
- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) = P(A,B)$ is the **joint probability** of events A and B. Meaning both events are true.
- $P(A|B) = P(A \cap B)/P(B)$ is the **conditional probability** of the event A being true given the event B is true.
- $P(A) = \sum_B P(A,B) = \sum_B P(A|B)P(B)$ is the **marginal probability** of the event A.

## 2.1 Likelihood

Most of the time when you are trying to model something, you have two things:

- Observations/data $\mathbf{x}$.
- A probabilistic model $P(x|\theta)$

And your goal is to estimate the hidden properties $\theta$ that gave rise to the data $\mathbf{x}$.

For example, if your probabilistic model is a Gaussian distribution, $\mathcal{N}(x_i,\mu,\sigma)$ your goal is to find the parameters $\theta = \{\mu,\sigma\}$ that maximize the probability that your data $\mathbf{x}$ were obtained given those parameters.

A classical method to achieve such result is the **maximum likelihood**, which consists of maximize the probability of the model with regard to $\theta$.

$$\hat{\theta} = argmax\, P(x|\theta)$$

This equation translates the fact the that you want to find the parameters $\theta$ that maximize the probability that your data $\mathbf{x}$ were indeed obtained by your probabilistic model.

### 2.2 Bayesian Inference

For Bayesian inference, we do not focus on the likelihood function $L(y) = P(x|y)$, but instead focus on the posterior distribution:

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

which is composed of the **likelihood** function $P(x|y)$, the **prior** $P(y)$ and a normalizing term $P(x)$

## 3 TP2 - Bayesian inference with one measurement

The Bayes theorem is the fundamental element of Bayesian inference. It states that:

### 3.1

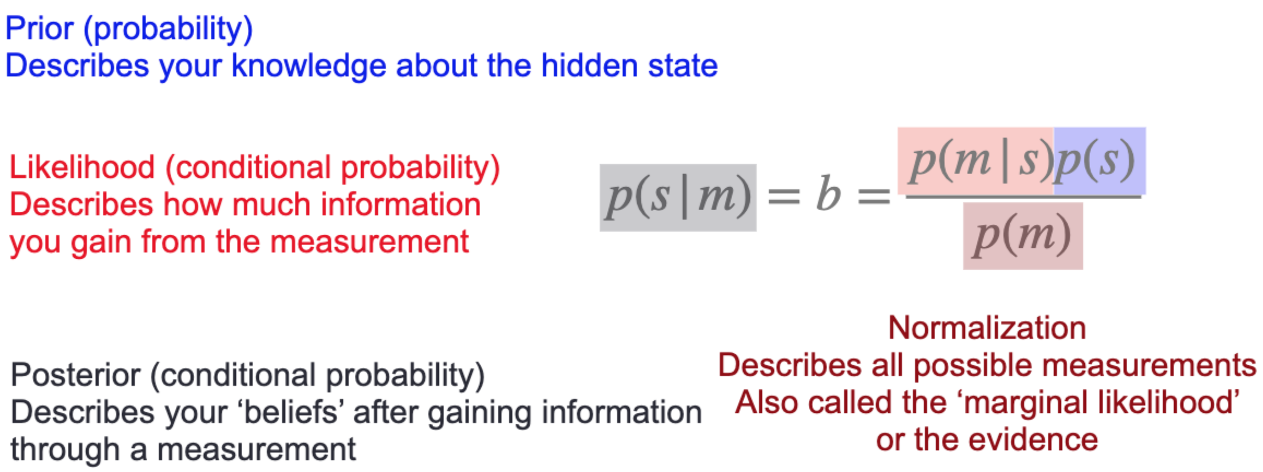



Figure 1: Bayes' rule

Marginalization is going to be used to combine our prior knowledge, which we call the **prior**, and our new information from a measurement, the **likelihood**. Only in this case, the information we gain about the hidden state we are interested in, where the fish are, is based on the relationship between the probabilities of the measurement and our prior.

We can now calculate the full posterior distribution for the hidden state ($s$) using Bayes' Rule. As we've seen, the posterior is proportional the prior times the likelihood. This means that the posterior probability of the hidden state ($s$) given a measurement ($m$) is proportional to the likelihood of the measurement given the state times the prior probability of that state:

$$P(s|m) \propto P(m|s)P(s)$$

We say proportional to instead of equal because we need to normalize to produce a full probability distribution:

$$P(s|m) = \frac{P(m|s)P(s)}{P(m)}$$

Normalizing by this $P(m)$ means that our posterior is a complete probability distribution that sums or integrates to 1 appropriately. We now can use this new, complete probability distribution for any future inference or decisions we like! In fact, as we will see tomorrow, we can use it as a new prior! Finally, we often call this probability distribution our beliefs over the hidden states, to emphasize that it is our subjective knowledge about the hidden state.

For many complicated cases, like those we might be using to model behavioural or brain inferences, the normalization term can be intractable or extremely complex to calculate. We can be careful to choose probability distributions where we can analytically calculate the posterior probability or numerical approximation is reliable. Better yet, we sometimes don't need to bother with this normalization! The normalization term, $P(m)$, is the probability of the measurement. This does not depend on state, so is essentially a constant we can often ignore. We can compare the unnormalized posterior distribution values for different states because how they relate to each other is unchanged when divided by the same constant. We will see how to do this to compare evidence for different hypotheses tomorrow. (It's also used to compare the likelihood of models fit using maximum likelihood estimation)

In this relatively simple example, we can compute the marginal likelihood $P(m)$ easily by using:

$$P(m) = \sum_s P(m|s)P(s)$$

We can then normalize so that we deal with the full posterior distribution.

## 3.2 Utility

We quantify gains and losses numerically using a **utility** function $U(s, a)$, which describes the consequences of your actions: how much value you gain (or, if negative, lose) given the state of the world ($s$) and the action you take ($a$). You need to calculate the expected utility by weighing the utilities with the probability of that state occurring. This allows us to choose actions by taking probabilities of events into account: we don't care if the outcome of an action-state pair is a loss if the probability of that state is very low. We can formalize this as:

$$\text{Expected utility of action a} = \sum_s U(s, a)P(s)$$

# 4 TP3 - Bayesian inference of a continuous hidden state

Bayes' rule tells us how to combine two sources of information: the prior (e.g., a noisy representation of Ground Control's expectations about where Astrocat is) and the likelihood (e.g., a noisy representation of the Astrocat after taking a measurement), to obtain a posterior distribution (our belief distribution) taking into account both pieces of information. Remember Bayes' rule:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Normalization constant}} \tag{1}$$

We will look at what happens when both the prior and likelihood are Gaussians. In these equations, $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with parameters $\mu$ and $\sigma^2$:

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

When both the prior and likelihood are Gaussians, Bayes Rule translates into the following form[1]:

[1] Prove the multiplying Gaussians for your self if you want

$$
\begin{aligned}
\text{Likelihood} &= \mathcal{N}(\mu_{likelihood}, \sigma^2_{likelihood}) \\
\text{Prior} &= \mathcal{N}(\mu_{prior}, \sigma^2_{prior}) \\
\text{Posterior} &= \mathcal{N}\left(\frac{\sigma^2_{likelihood}\mu_{prior} + \sigma^2_{prior}\mu_{likelihood}}{\sigma^2_{likelihood} + \sigma^2_{prior}}, \frac{\sigma^2_{likelihood}\sigma^2_{prior}}{\sigma^2_{likelihood} + \sigma^2_{prior}}\right) \\
&\propto \mathcal{N}(\mu_{likelihood}, \sigma^2_{likelihood}) \times \mathcal{N}(\mu_{prior}, \sigma^2_{prior})
\end{aligned}
$$

# 5 TP4 - Hidden Markov model

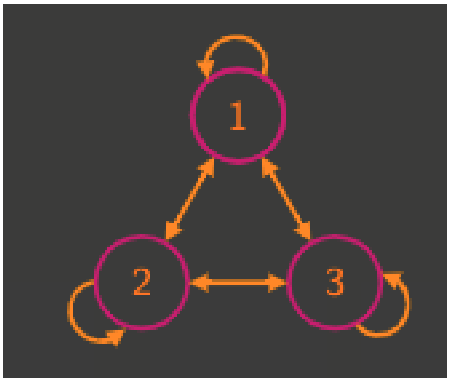



Figure 2: Markov chains

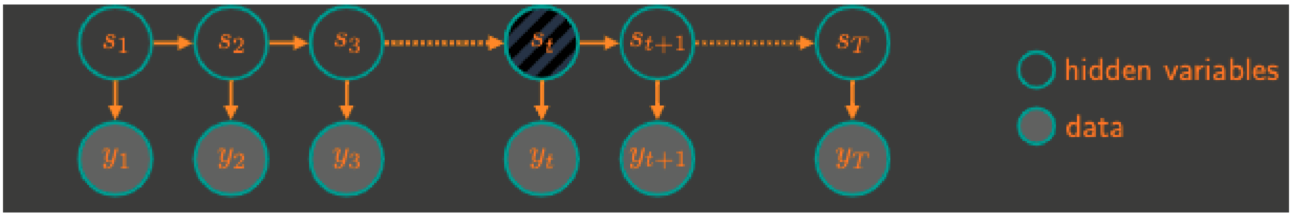



Figure 3: HMM graphical model

In a hidden Markov model, the future is **independent** of the past given the present, it is memoryless.

**the transition matrix:**

$$D = \begin{pmatrix} p_{\text{stay}} & p_{\text{switch}} \\ p_{\text{switch}} & p_{\text{stay}} \end{pmatrix}$$

with $p_{\text{stay}} = 1 - p_{\text{switch}}$.

*If state could be continuous ($s_t \in \mathbb{R}^k$), we will use the Kalman filter.*

**example with two states** For two states, We can summarize these as a $2 \times 2$ matrix we will denote $D$ for Dynamics.

$$D = \begin{bmatrix} p(s_t = +1 | s_{t-1} = +1) & p(s_t = -1 | s_{t-1} = +1) \\ p(s_t = +1 | s_{t-1} = -1) & p(s_t = -1 | s_{t-1} = -1) \end{bmatrix}$$

$D_{ij}$ represents the transition probability to switch from state $i$ to state $j$ at the next time step. We can represent the probability of the $_{c}urrent_{s}tateasa2 - dimensionalvector$
$P_t = [p(s_t = +1), p(s_t = -1)]$

. The entries are the probability that the current state is +1 and the probability that the current state is -1 so these must sum up to 1.

We then update the probabilities over time following the Markov process:

$$P_t = P_{t-1} D \tag{1}$$

If you know the state, the entries of $P_{t-1}$ would be either 1 or 0 as there is no uncertainty.

**Measurements:** In a *Hidden* Markov model, we cannot directly observe the latent states $s_t$. Instead, we get noisy measurements $m_t \sim p(m|s_t)$.

### 5.1 Forward inference of HMM

As a recursive algorithm, let's assume we already have yesterday's posterior from time $t-1$: $p(s_{t-1}|m_{1:t-1})$. When the new data $m_t$ comes in, the algorithm performs the following steps:

- **Predict**: transform yesterday's posterior over $s_{t-1}$ into today's prior over $s_t$ using the transition matrix $D$:

$$\text{today's prior} = p(s_t|m_{1:t-1}) = p(s_{t-1}|m_{1:t-1})D$$

- **Update**: Incorporate measurement $m_t$ to calculate the posterior $p(s_t|m_{0:t})$

$$\text{posterior} \propto \text{prior} \cdot \text{likelihood} = p(m_t|s_t)p(s_t|m_{0:t-1})$$

## 6 TP5 - The Kalman Filter

Like in Hidden Markov Models, one state only depends on the previous state, the past history does not matter.

In one dimension, you have the following relationship:

$$x[t+1] \quad = \quad ax[t] + \xi[t] \tag{2}$$

Where $a$ is a deterministic known parameter and $\xi$ is Gaussian motor noise generated from $\mathcal{N}(0, \sigma^2)$.

As we saw in the previous tutorials, our sensory inputs are not perfect, and it is impossible to know the real latent state; we can only get a rough estimate of it. Mathematically, we define the observation of the latent state of our dynamical system, $y[t]$ as follows:

$$\begin{aligned} y[t] &= Hx[t] + \eta[t] && (3) \\ \begin{bmatrix} y_1[t] \\ y_2[t] \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1[t] \\ x_2[t] \end{bmatrix} + \begin{bmatrix} \eta_1[t] \\ \eta_2[t] \end{bmatrix} && (4) \end{aligned}$$

where $H$ is the called the *observation matrix* and $\eta[t]$ is a vector of Gaussian white sensory noise generated from $\mathcal{N}(0, \Omega_{\text{sensory}})$.

## 6.1 The Kalman Filter, or the optimal estimation of continuous linear dynamical systems

**Step 1: Change yesterday's posterior into today's prior**

The first step consists into updating yesterday's posterior into today's prior while taking into account the system dynamics. Recall that yesterday's posterior is a Gaussian distribution $\mathcal{N}(\mu_{s_{t-1}}, \sigma^2_{s_{t-1}})$. The mathematical model of the system is composed of a deterministic shift $a$ and some additional noise. When you multiply your prior by $a$, you multiply each point of the distribution. Therefore, you will get a new distribution:

$$\mathcal{N}(a * \mu_{s_{t-1}}, a^2 * \sigma^2_{s_{t-1}})$$

Then, you add the process noise and get:

$$\mathcal{N}(a * \mu_{s_{t-1}}, a^2 * \sigma^2_{s_{t-1}} + \sigma^2_{motor})$$

You have now today's prior !

**Step 2: Multiply today's prior by likelihood**

Use the latest measurement to form a new estimate somewhere between this measurement and what we predicted in Step 1. The next posterior is the result of multiplying the Gaussian computed in Step 1 (a.k.a. today's prior) and the likelihood, which is also modelled as a Gaussian $\mathcal{N}(m_t, \sigma^2_{sensory})$:

**2a: add information from prior and likelihood**

To find the posterior variance, we first compute the posterior information (which is the inverse of the variance) by adding the information provided by the prior and the likelihood:

$$\frac{1}{\sigma^2_{s_t}} = \frac{1}{a^2 * \sigma^2_{s_{t-1}} + \sigma^2_{motor}} + \frac{1}{\sigma^2_{sensory}}$$

Now we can take the inverse of the posterior information to get back the posterior variance.

**2b: add means from prior and likelihood**

To find the posterior mean, we calculate a weighted average of means from prior and likelihood, where each weight, $g$, is just the fraction of information that each Gaussian provides!

$$g_{\text{prior}} = \frac{\text{information}_{\text{ prior}}}{\text{information}_{\text{ posterior}}}$$

$$g_{\text{likelihood}} = \frac{\text{information}_{\text{ likelihood}}}{\text{information}_{\text{ posterior}}}$$

$$\bar{\mu}_t = g_{\text{prior}} * a * \mu_{s_{t-1}} + g_{\text{likelihood}} * m_t$$

## 6.2 Relationship to classic description of Kalman filter

The above weights, $g_{\text{prior}}$ and $g_{\text{likelihood}}$, add up to 1 and can be written one in terms of the other; then, if we let $K = g_{\text{likelihood}}$, the posterior mean can be expressed as:

$$\bar{\mu}_t = (1 - K)D\bar{\mu}_{t-1} + Km_t = D\bar{\mu}_{t-1} + K(m_t - D\bar{\mu}_{t-1})$$

In classic textbooks, you will often find this expression for the posterior mean; $K$ it is known as the Kalman gain and its function is to choose a value partway between the current measurement $m_t$ and the prediction from Step 1.

# 7 TP6 - Optimal feedback control - LQG