EPFL

NX-414
A. Mathis
M. Schrimpf

Week 3 - Exercises on Bayesian Perception - Solutions

The brain needs to integrate multiple sensory streams to obtain a representation about of the world. As we saw in today's lecture, a typical example is the integration of haptic and visual feedback. Furthermore, inference is also important for motor control and learning. Bayesian statistics offers a theoretical framework to interpret how humans and animals solve the sensory integration problem. Overall, the goal of today's exercise is to:

- Learn to derive step-by-step the rule used by a decision making system to fuse different sensory information, according to the Bayesian interpretation.
- Compare the Bayesian and the maximum likelihood interpretation of sensory integration.
- Find the minimum variance weighted average of the sensory inputs.

Exercise 1.1 - Integration of two sensory channels

Let us consider the two Gaussian random variables t and v, corresponding to the (noisy) estimation of the width w of an object through tactile and visual feedback. We assume that the two variables are independent, and that they are distributed according to the following conditional density function:

$$\begin{aligned} p(t|w) &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(w-t)^2}{2\sigma_t^2}\right) \\ p(v|w) &= \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(w-v)^2}{2\sigma_v^2}\right) \end{aligned} \tag{1}$$

Show that, assuming a flat prior distribution $p(w)$, the posterior is proportional to the product of the likelihoods, i.e.,

$$p(w|t,v) \propto p(t|w)p(v|w) \tag{2}$$

Solution:

$$p(w|t,v) = \frac{p(t,v|w)p(w)}{p(t,v)} = \frac{p(t|w)p(v|w)p(w)}{p(t,v)} \propto p(t|w)p(v|w) \tag{3}$$

In the first step we used the Bayes rule, in the second the independence of t and v, in the third that $p(w)$ is uniform and that $p(t,v)$ does not depend on w.

Exercise 1.2

Re-write the analytical expression of $p(t|w)p(v|w)$ to show that it is the density function of a Gaussian distribution of variable w (up to a normalization coefficient independent of w).

Solution: Simply complete the square:

$$
\begin{aligned}
p(t|w)p(v|w) &= C \exp\left(-\frac{(w-t)^2}{2\sigma_t^2} - \frac{(w-v)^2}{2\sigma_v^2}\right) \\
&= C \exp\left(-\frac{(\sigma_v^2+\sigma_t^2)w^2 - 2(\sigma_v^2 t + \sigma_t^2 v)w + \sigma_v^2 t^2 + \sigma_t^2 v^2}{2\sigma_t^2\sigma_v^2}\right) \\
&= C \exp\left(-\frac{w^2 - 2\frac{\sigma_v^2 t + \sigma_t^2 v}{\sigma_v^2+\sigma_t^2} w + K}{2\frac{\sigma_t^2\sigma_v^2}{\sigma_t^2+\sigma_v^2}}\right)
\end{aligned}
\tag{4}
$$

Here, C and K indicate two terms not depending on w. The values of the two constants must be such that the numerator of the exponential is a square and that the integral of the density amounts to one, but their actual expression is irrelevant for the rest of the exercise.

Exercise 1.3

Given the density function derived in Exercise 1.2, extract the mean μ_w and the variance σ_w^2 of w as a function of t, v, σ_t and σ_v, and show that μ_w is the weighted average of t and v, with weights proportional to the inverse of their variance.

Solution:

In the expression found in Exercise 1.2 we just need the coefficient of w at the numerator of the exponential to find μ_w and the denominator to find σ_w^2. In fact, in a Gaussian distribution, we have:

$$
p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2 - 2\mu x + \mu^2}{2\sigma^2}\right) \tag{5}
$$

It follows that:

$$
-2\mu_w = -2\frac{\sigma_v^2 t + \sigma_t^2 v}{\sigma_v^2 + \sigma_t^2} \tag{6}
$$

and

$$
2\sigma_w^2 = 2\frac{\sigma_t^2\sigma_v^2}{\sigma_t^2 + \sigma_v^2} \tag{7}
$$

from which we have

$$
\mu_w = \frac{\sigma_v^2 t + \sigma_t^2 v}{\sigma_v^2 + \sigma_t^2} = \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_t^2}\right)^{-1} \left(\frac{1}{\sigma_t^2} t + \frac{1}{\sigma_v^2} v\right) \tag{8}
$$

and

$$
\sigma_w^2 = \frac{\sigma_t^2\sigma_v^2}{\sigma_t^2 + \sigma_v^2} = \left(\frac{1}{\sigma_v^2} + \frac{1}{\sigma_t^2}\right)^{-1} \tag{9}
$$

Exercise 1.4

Show that the variance of W is smaller than both the variance of V and of the variance of T, for any value of $\sigma_t > 0$ and $\sigma_v > 0$.

Solution: We show this fact for σ_t, but the procedure is identical for σ_v. We write the claim and we prove that it can be reduced to a true inequality:

$$\sigma_t^2 > \frac{\sigma_v^2 \sigma_t^2}{\sigma_t^2 + \sigma_v^2} \Leftrightarrow \quad \frac{\sigma_t^4 + \sigma_t^2\sigma_v^2 - \sigma_t^2\sigma_v^2}{\sigma_t^2 + \sigma_v^2} > 0 \Leftrightarrow \sigma_t^4 > 0 \Leftrightarrow \quad \sigma_t \neq 0 \tag{10}$$

which is true by assumption.

Exercise 2.1 - Extension to multiple sensory channels

Consider a new estimator u of w, through another sensory input (or sensor), independent of v and t, and also following a Gaussian distribution:

$$p(u|w) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{(w-u)^2}{2\sigma_u^2}\right) \tag{11}$$

Write the new expression of μ_w and σ_w^2.

Solution: Without repeating all the computation, we observe that the old estimation of $\mu_{\tilde{w}}$ is also distributed as a Gaussian, with mean and variance derived in Exercise 1.3. Therefore, the new variance of the estimator is given by:

$$\sigma_w^2 = \left(\frac{1}{\sigma_{\tilde{w}}^2} + \frac{1}{\sigma_u^2}\right)^{-1} = \left(\frac{1}{\sigma_t^2} + \frac{1}{\sigma_v^2} + \frac{1}{\sigma_u^2}\right)^{-1} \tag{12}$$

Similarly,

$$\begin{aligned}
\mu_w &= \left(\frac{1}{\sigma_{\tilde{w}}^2} + \frac{1}{\sigma_u^2}\right)^{-1} \left(\frac{1}{\sigma_{\tilde{w}}^2}\mu_{\tilde{w}} + \frac{1}{\sigma_u^2}u\right) \\
&= \left(\frac{1}{\sigma_t^2} + \frac{1}{\sigma_v^2} + \frac{1}{\sigma_u^2}\right)^{-1} \left(\left(\frac{1}{\sigma_t^2} + \frac{1}{\sigma_v^2}\right)\left(\frac{1}{\sigma_t^2} + \frac{1}{\sigma_v^2}\right)^{-1}\left(\frac{1}{\sigma_t^2}t + \frac{1}{\sigma_v^2}v\right) + \frac{1}{\sigma_u^2}u\right) \\
&= \left(\frac{1}{\sigma_t^2} + \frac{1}{\sigma_v^2} + \frac{1}{\sigma_u^2}\right)^{-1} \left(\frac{1}{\sigma_t^2}t + \frac{1}{\sigma_v^2}v + \frac{1}{\sigma_u^2}u\right)
\end{aligned} \tag{13}$$

Exercise 2.2

Provide the expression of μ_w and σ_w^2 given an arbitrary number N of independent Gaussian estimators $t_1, ..., t_N$, normally distributed with mean w and variance σ_i, $i = 1, ..., N$.

Solution: The proof of the induction step is the same as the proof of 2.1. Without rewriting all the details, we can conclude that N estimators lead to the following expressions:

$$\sigma_w^2 = \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-1} \quad \text{and} \quad \mu_w = \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right)^{-1} \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} t_i\right) \tag{14}$$

Exercise 3.1 - Maximum likelihood approach

Let us now consider the problem of multisensory integration from a different point of view. Instead of the Bayesian approach, we want to estimate w through maximum likelihood. Consider the conditional distributions $p(v|w)$ and $p(t|w)$ as before and compute the maximum likelihood estimator for μ_w. Derive the variance of such estimator and show that you retrieve the same result as in Exercise 1.3. Can the variance of an unbiased estimator of μ_w be lower than the value you have found?

Solution:

$$\begin{aligned}
&\mathcal{L}(w) = p(t|w)p(v|w) = C\exp\left(-\frac{(t-w)^2}{2\sigma_t^2} - \frac{(v-w)^2}{2\sigma_v^2}\right)\\
&\frac{d}{dw}\mathcal{L}(w) = C\left(-\frac{t-w}{\sigma_t^2} - \frac{v-w}{\sigma_v^2}\right)\exp\left(-\frac{(t-w)^2}{2\sigma_t^2} - \frac{(v-w)^2}{2\sigma_v^2}\right) = 0 \Leftrightarrow\\
&-\sigma_v^2 t + \sigma_v^2 w - \sigma_t^2 v + \sigma_t^2 w = 0 \Leftrightarrow\\
&w = \frac{\sigma_v^2}{\sigma_t^2+\sigma_v^2}t + \frac{\sigma_t^2}{\sigma_t^2+\sigma_v^2}v
\end{aligned} \tag{15}$$

With simple manipulation it is possible to show that we have obtained the same estimator for the parameter w, and its variance corresponds to σ_w as found before. It is the minimum variance for an unbiased estimator, as it is at the Cramer-Rao limit (you can verify that the inverse of the variance is the sum of the Fisher information of v and t, as shown in the lecture).

Exercise 3.2

In the derivation through maximum likelihood, we have implicitly assumed that all the information about w is provided by v and t. How was this hypothesis included in the Bayesian multisensory integration?

Solution: In the Bayesian formulation we had assumed a flat prior for w, which is equivalent to assigning an equal probabilty to every value.

Exercise 4.1 - Minimum variance weighted average

We consider once again the problem of sensory integration from a different point of view. Let us consider a weighted sum of noisy "measurements" of w, defined as $W = \sum_{i=1}^N \alpha_i T_i$, in which the index i indicates a sensory channel. The measurements t_i are independent random variables for T_i with mean w and variance σ_i^2, $i = 1, ..., N$. First, show that the necessary and sufficient condition for W to be unbiased is that $\sum_i^N \alpha_i = 1$. Second, use the technique of the Lagrange multiplier to show that the coefficients α_i are the ones found in Exercise 2.2.

Solution: First we show that $\sum_{i=1}^N \alpha_i = 1$. We just need to compute the expected value of W:

$$\mathbb{E}_w[W] = \mathbb{E}[\sum_{i=1}^N \alpha_i T_i] = \sum_{i=1}^N \alpha_i \mathbb{E}[T_i] = w\sum_{i=1}^N \alpha_i \tag{16}$$

which is equal to w if and only if $\sum_{i=1}^{N} \alpha_i = 1$. We can therefore cast the minimization problem as:

$$\text{Minimize} \quad \mathbb{V}[W] = \sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 \qquad \text{subject to} \quad \sum_{i=1}^{N} \alpha_i = 1 \tag{17}$$

This problem can be solved minimizing the Lagrangian

$$\mathcal{L}(\alpha_0, ..., \alpha_N) = \sum_{i=1}^{N} \alpha_i^2 \sigma_i^2 - \alpha_0 \left(\sum_{i=1}^{N} \alpha_i - 1 \right) \tag{18}$$

We compute the partial derivatives w.r.t. α_k, $k = 1, ..., N$:

$$\frac{\partial}{\partial \alpha_k} \mathcal{L} = 2\alpha_k \sigma_k^2 - \alpha_0 = 0 \tag{19}$$

which implies that $\alpha_k = \frac{\alpha_0}{2\sigma_k^2}$. As we have that $\sum_{i=1}^{N} \alpha_i = 1$:

$$\sum_{i=1}^{N} \frac{\alpha_0}{2\sigma_i^2} = 1 \Rightarrow \alpha_0 = 2 \left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \right)^{-1} \Rightarrow \alpha_k = \left(\sum_{i=1}^{N} \frac{1}{\sigma_i} \right)^{-1} \frac{1}{\sigma_k^2} \tag{20}$$