Building blocks of Bayesian models – Likelihoods, Priors and Posteriors

Sensor fusion & nonlinear filtering

Lars Hammarstrand

LIKELIHOODS, PRIORS AND POSTERIORS

General problem formulation

- We are interested in an unknown parameter $\theta \in \Theta$ using some related observed data y.
- Common problem types are estimation (e.g., $\Theta = \mathbb{R}^n$) and detection problems (e.g., $\Theta = \{-1, 1\}$).

Assumption

- The observed data, y, is distributed as

$$y \sim p(y|\theta),$$

where p is a known distribution.

LIKELIHOODS, PRIORS AND POSTERIORS

Likelihood

- Since y is observed, we often view $p(y|\theta)$ as a function of θ,

$$l(\theta|y) = p(y|\theta),$$

  where $l(\theta|y)$ is called the likelihood function.

- Note: the likelihood function is *not* a density w.r.t. θ.

Prior

- In Bayesian statistics we have a prior distribution on θ, $p(\theta)$.
- Prior means *earlier*, or before, and $p(\theta)$ describes what we know *before* observing y.

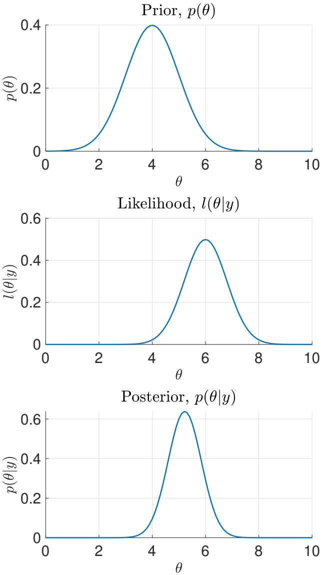
LIKELIHOODS, PRIORS AND POSTERIORS

Posterior

- One objective in Bayesian statistics is to compute the posterior

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto l(\theta|y)p(\theta)$$

- Posterior means *after* and $p(\theta|y)$ describes what we know *after observing y*.

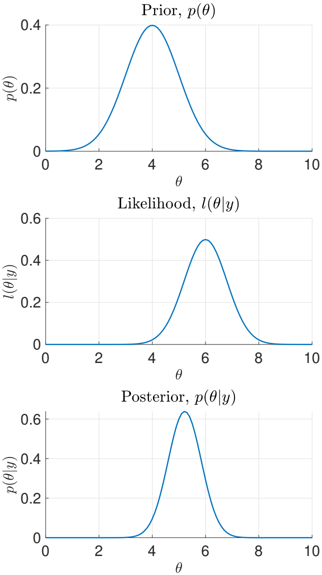


LIKELIHOODS, PRIORS AND POSTERIORS

- We summarize this as

$$\text{posterior} \propto \text{likelihood} \times \text{prior}.$$

- Given the posterior, $p(\theta|y)$ we can answer, e.g.,
  - What is the most probable θ?
  - What is the probability that $\theta \in \mathcal{A}$?
  - What is the posterior mean of θ?
- We can also minimize expected costs in a decision theoretic manner.



EXAMPLE: SCALAR IN GAUSSIAN NOISE

Estimation of scalar in Gaussian noise

- Suppose we observe

$$y = \theta + v, \qquad v \sim \mathcal{N}(0, \sigma^2)$$

  such that $p(y \big| \theta) = \mathcal{N}(y; \theta, \sigma^2) \propto \exp\{-(y - \theta)^2 / (2\sigma^2)\}$.

- A common non-informative prior on θ is $p(\theta) \propto 1$.

- What is the posterior?

BAYESIAN APPROACH TO SENSOR FUSION

- Suppose we collect measurements from two types of sensors, y_1 and y_2,

Bayesian fusion of independent observations

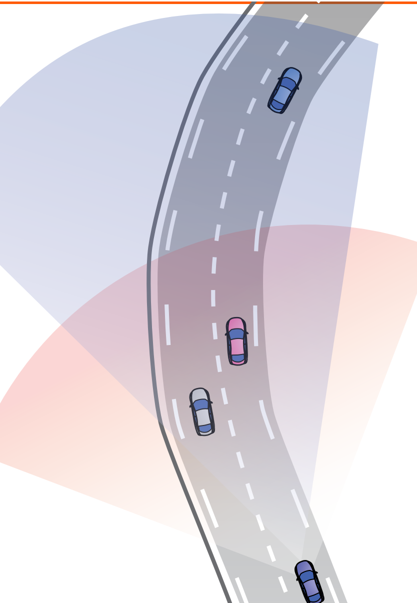
- We seek the posterior distribution:

$$p(\theta|y_1, y_2) \propto p(\theta)p(y_1, y_2|\theta).$$

- It is often reasonable to assume that

$$p(y_1, y_2|\theta) \approx p(y_1|\theta)p(y_2|\theta),$$

i.e., that measurements are conditionally independent.



SELF-ASSESSMENT

The posterior distribution is $p(\theta \middle| y) \propto p(y \middle| \theta) p(\theta)$.
It is also true that:

- The normalization factor is not always unique?
- The posterior $p(\theta \middle| y)$ can always be uniquely determined from the fact that $\int p(\theta \middle| y)\, d\theta = 1$?
- The posterior distribution can only be uniquely determined if it is proportional to a well known distribution, e.g., a Gaussian.

Only one statement is correct.