Bayesian Decision Theory

Sensor fusion & nonlinear filtering

Lars Hammarstrand

BAYESIAN DECISION PRINCIPLE

- How can we use $p(\theta|y)$ to make decisions?
- **Examples** of decision problems
  - How to control a self-driving vehicle.
  - How to invest money.
  - Select medicine to give to a patient
  - Estimate a parameter vector (may represent temperature, distance, etc).

Basic principle of Bayesian decision theory

- Minimize expected loss

  or, equivalently,
- Maximize expected utility.

DECISION THEORY – A TOY EXAMPLE

Choosing a course

- A student wants to decide whether to take a course or not.
- Suppose $\theta \in \{$good course, fair course, bad course$\}$ and

	good course	fair course	bad course
$\Pr\{\theta \mid y\}$	0.3	0.3	0.4

- If the loss function is

	good course	fair course	bad course
Take	0	5	30
Not take	20	5	0

should he/she then take the course?

Taking the course

$0.3 \times 0 + 0.3 \cdot 5 + 0.4 \times 30$

$= 10.5$

Not taking the course

$0.3 \times 20 + 0.3 \times 5 + 0.4 \times 0$

$= 7.5$

MINIMUM POSTERIOR EXPECTED LOSS

- We often study loss functions $C(\theta, a)$ instead of utility. (Typically, $C \geq 0$.)
- Let $\hat{\theta}$ denote an estimate of θ.

Optimal Bayesian decisions

Minimize the posterior expected loss

$$\hat{\theta} = \arg\min_{a} \mathbb{E}\left\{C(\theta, a)\middle|y\right\}$$

where $\mathbb{E}\left\{C(\theta, a)\middle|y\right\} = \int_{\Theta} C(\theta, a)p(\theta|y)\,d\theta$

- Note: y is given (fixed) and θ is random.

SELF-ASSESSMENT

To make an optimal Bayesian decision it is sufficient to know:

- The prior, $p(\theta)$, the likelihood, $p(y|\theta)$, and a loss function $C(\theta, a)$.
- The likelihood, $p(y|\theta)$, and a loss function $C(\theta, a)$.
- The posterior distribution, $p(\theta|y)$, and a loss function, $C(\theta, a)$.

Check all statements that apply.

COMPARISON: BAYES VS FREQUENTIST

Frequentist	Bayes
θ is fixed and unknown $\Rightarrow$ θ is deterministic	Uncertainties in θ are described stochastically $\Rightarrow$ θ is random
Maximum likelihood (ML) most famous estimator $\hat{\theta}_{ML} = \arg\max_\theta l(\theta \vert y)$	Minimum mean square error and maximum a posteriori estimators, e.g., $\hat{\theta}_{MAP} = \arg\max_\theta p(\theta) l(\theta \vert y)$
Study performance by averaging over y for fixed θ	Make decisions conditioned on the observation y.

- Note 1: most Bayesians also study frequentist performance.
- Note 2: many frequentists agree that parameters may be random in some situations.