28/04/2022. LECTURE 21: NOTES MADE ON THE VISUALISER.

In practice, for the M-H algorithm we want:

- A high acceptance probability
  don't want to repeatedly sample points we reject
  good $q(\phi|\theta)$ close to the target but slightly heavier in the tails
- To explore the whole distribution
  good coverage of all possible values of θ. Don't want to spend too long in only a small area of the distribution.
  chain that does not move around quickly is "slow-mixing"

The Gibbs sampler can experience slow-mixing if the parameters are highly correlated. Ideally, we want the parameters to be close to independence.

4. DECISION THEORY.

Any situation in which choices must be made amongst two or more possible alternate courses of action is a decision problem.

- If we are certain of the consequences of these choices (or actions) then the decision is relatively simple. You can write down all the possible options and choose the decision you like the best.
- When the consequences are uncertain then the problem is much less straightforward (and therefore much more interesting!)
- When study how decisions ought to be made. In short, what is the optimal decision?

Consider an experiment with possible outcomes $\theta \in \Theta$ [" true state of nature"] Without knowing θ, we must make a decision d from a class D of possible decisions when the consequences of the choice of d depend upon θ.

[e.g. θ could be performance in the exam, the decision whether or not to do a specified problem sheet].

Having chosen d we observe θ. We need a way of assessing how good or bad this choice is. We quantify this by a loss function $L(\theta, d)$ which specifies how bad it would be to make decision d if the outcome was θ.

i.e. we prefer d_1 to d_2 if $L(\theta, d_1) < L(\theta, d_2)$

Thus, we want to minimise loss.

[A related measure is a utility function $U(\theta, d)$ which works with goodness.

- The ideas of utility and how to construct a utility function to compare anything is covered in Section 4.1 of the on-line notes.
- We won't cover this and you won't be assessed on anything in Section 4.1.
- Often $U(\theta, d) = -L(\theta, d)$]

We'll focus on statistical decision theory when we consider inference about a parameter θ as a decision problem.

There are three main types of inference and each of these can be addressed as a decision problem by a suitable choice of D.

Point Estimation $d \in \Theta$ (Point)

Set/Interval Estimation $d \subset \Theta$ (Subset) [2019/20 exam, Q4(c) $d = (\theta_1, \theta_2)$ when $\theta_1 \leq \theta_2$]

Hypothesis test d member of a partition of Θ

(e.g. $\Theta = \mathbb{R}$, $H_0: \theta < 0$, $H_1: \theta \geq 0$)

In examples, we'll typically focus on point estimation and so specify a single value d as an estimate of θ.

A statistical decision has a number of parts or ingredients

- The possible values of the parameter θ, Θ THE PARAMETER SPACE.
- The set of possible decisions, $\mathcal{D}$ THE DECISION SPACE
- The probability distribution on θ, $\pi(\theta)$.
  - prior $\pi(\theta) = f(\theta)$ IMMEDIATE DECISION
  - posterior given data x, $\pi(\theta) = f(\theta|x)$ DECISION AFTER OBSERVING x
  - posterior given data x, y $\pi(\theta) = f(\theta|x,y)$ DECISION AFTER OBSERVING x, y
- The LOSS FUNCTION $L(\theta, d)$

The decision problem is completely specified by $[\Theta, \mathcal{D}, \pi(\theta), L(\theta, d)]$

The RISK of d under $\pi(\theta)$ is the expected loss

$$\rho(\pi, d) = \underbrace{\int_\theta \overbrace{L(\theta, d)}^{\text{function of } \theta \text{ and } d} \pi(\theta)\, d\theta}_{\text{function of } d}$$

Thus, for each $d \in \mathcal{D}$ I have the corresponding risk $\rho(\pi, d)$. I can choose d to minimise this.

DEFINITION (Bayes rule and risk).

The Bayes risk $\rho^*(\pi)$ minimises the expected loss

$$\rho^*(\pi) = \inf_{d \in \mathcal{D}} \rho(\pi, d)$$

with respect to $\pi(\theta)$. A decision $d^* \in \mathcal{D}$ for which $\rho(d^*, \pi) = \rho^*(\pi)$ is a Bayes (decision) rule against $\pi(\theta)$.

The Bayes rule may not be unique and, in weird cases, might not exist

Typically we solve $[\Theta, \mathcal{D}, \pi(\theta), L(\theta, d)]$ by finding $\rho^*(\pi)$ and (at least one) d^*

EXAMPLE (Quadratic loss)

$$L(\theta, d) = (\theta - d)^2$$

[For quadratic loss, under $\pi(\theta)$, the Bayes rule is $\mathbb{E}_{(\pi)}(\theta)$ and the Bayes risk is $\mathrm{Var}_{(\pi)}(\theta)$].