STAT40810 — Stochastic Models

Brendan Murphy

Week 3

Wrong Model

What happens if we fit the wrong model?

- When constructing a statistical model, we usually make a number of assumptions.
- One may ask "What happens if we fit the wrong model?"
- We would like to feel that inferences are somewhat robust to modeling choice.
- I will demonstrate some weak form of robustness to model choice.
- This offers a modicum of comfort when choosing a model.

Preliminary: Kullback-Leibler Divergence

- A useful measure of difference between two probability distributions is Kullback-Leibler divergence.
- It is defined as follows:

$$
\begin{aligned}
D(f||g) &= \int f(x) \log \frac{f(x)}{g(x)} dx \text{ when continuous} \\
D(f||g) &= \sum f(x) \log \frac{f(x)}{g(x)} dx \text{ when discrete}
\end{aligned}
$$

- It is worth noting that:
  - $D(f||g) \geq 0$
  - $D(f||g) = 0$ if and only if $f = g$.

Maximum Likelihood

- Suppose that we have $x_1, x_2, \ldots, x_n$ which are sampled from some model $g(x)$.
- But we model the data as coming from $f(x|\theta)$ with unknown θ.
- Let's assume that we have continuous values (for simplicity)
- When fitting a model, we may use maximum likelihood.
- When doing this we find a value of θ that maximizes

$$\ell(\theta) = \sum_{i=1}^{n} \log f(x_i|\theta).$$

- Equivalently, we get a θ that minimizes

$$-\ell(\theta) = -\sum_{i=1}^{n} \log f(x_i|\theta).$$

Maximum Likelihood 2

- Equivalently, we get a θ that minimizes
$$-\frac{\ell(\theta)}{n} = -\frac{1}{n}\sum_{i=1}^{n} \log f(x_i|\theta).$$
- As $n \to \infty$, we get
$$-\frac{1}{n}\sum_{i=1}^{n} \log f(x_i|\theta) \to -\mathbb{E}\log f(x|\theta) = -\int g(x)\log f(x|\theta)dx$$
- So, as $n \to \infty$ we are minimizing
$$-\int g(x)\log f(x|\theta)dx$$
with respect to θ.
- This is equivalent to minimizing
$$\int g(x)\log g(x)dx - \int g(x)\log f(x|\theta)dx.$$

Maximum Likelihood 3

- So, we are effectively minimizing

$$
\begin{aligned}
D(g||f(x|\theta)) &= \int g(x)\log g(x)dx - \int g(x)\log f(x|\theta)dx \\
&= \int g(x)\log\frac{g(x)}{f(x|\theta)}dx
\end{aligned}
$$

- That is, maximum likelihood is (approximately) finding the closest model to the correct one.