Assignment 4

1 Biased coin

Suppose we flip a coin N times and observe N_0 tails and N_1 heads. We believe the outcome of a flip is determined by a Bernoulli with parameter θ.

1. Let $\hat{\theta}_1$ be the MAP under a uniform prior (i.e. $P(\theta) = 1$). Derive $\hat{\theta}_1$.

2. Consider the following prior for θ that believes the coin is either fair or slightly biased towards tails.

$$p(\theta) = \begin{cases} 0.5 & \text{if } \theta = 0.5 \\ 0.5 & \text{if } \theta = 0.4 \\ 0 & \text{otherwise} \end{cases}$$

   Derive the MAP estimate of θ under this prior, $\hat{\theta}_2$.

3. Suppose the true parameter is $\theta = 0.41$, and suppose $N = 10$. What is the probability that $\hat{\theta}_2$ is closer to the true value than $\hat{\theta}_1$? That is, what is

$$P(|\hat{\theta}_2 - \theta| < |\hat{\theta}_1 - \theta|)?$$

   You may find the following binomial distribution table for $\theta = 0.41$ helpful.

Binomial distribution (n=10, p=0.41)			
	f(x)	F(x)	1 - F(x)
x	Pr[$X = x$]	Pr[$X \leq x$]	Pr[$X > x$]
0	0.0051	0.0051	0.9949
1	0.0355	0.0406	0.9594
2	0.1111	0.1517	0.8483
3	0.2058	0.3575	0.6425
4	0.2503	0.6078	0.3922
5	0.2087	0.8166	0.1834
6	0.1209	0.9374	0.0626
7	0.0480	0.9854	0.0146
8	0.0125	0.9979	0.0021
9	0.0019	0.9999	0.0001
10	0.0001	1.0000	-0.0000

4. Suppose instead $N = 10000$. Which estimator do you think will usually be close to the true value? Why?

Assignment 4

2 Regularization vs prior

Let θ be a univariate, continuous parameter. Consider the following optimization, which adds a regularization term to the log likelihood to identify a regularized MLE.

$$\hat{\theta}_{reg} = \text{argmin}_\theta(-\log p(D|\theta) + \lambda C(\theta)) \qquad \lambda > 0$$

1. As an alternative, we can define the MAP estimate $\hat{\theta}_{MAP}$ for a particular choice of prior $p(\theta)$. Define a probability distribution $p(\theta)$ as a function of $C(\theta)$ such that $\hat{\theta}_{MAP} = \hat{\theta}_{reg}$. You may assume that the integral of C (and reasonable functions of C) are finite.
2. According to $p(\theta)$, what is the relative probability of two parameters $p(\theta_1)/p(\theta_2)$?
3. Qualitatively, how does $p(\theta)$ change as we decrease λ? Increase?

3 Survey

On average, how many hours have you spent on each assignment so far in this course?

What is one thing from this course so far that you found confusing, a topic you would like to hear more about, or something you found particularly interesting?