

EXAMPLE VIVA QUESTIONS

Preliminary question: Explain the analysis you did for homework 2 as if explaining to a general audience. What data did you use and what did you learn? [MAX 2-3 MINS]

1. What is the Law of Large Numbers? Prove it.
2. State the Central Limit Theorem. How can it be proved? (rough overview sufficient). Are the sample mean/variance unbiased estimators?
3. Suppose that an infected person enters a population, and infections then occur according to a random walk (with $P(\text{infection}) = 0.6$ and $P(\text{recovery}) = 0.4$). How can we estimate the probability that the pathogen fades out without causing a major epidemic?
4. What are the positive and negative predictive values for a test for infection? Calculate their respective values for the below infection table.

		True infection status	
		Infected	Uninfected
Test result	Positive	940/1000	0/1000
	Negative	50/1000	10/1000

5. How can uniformly sampled random numbers on $[0,1]$ be used to sample from a more complex probability density function?
6. What is the linear chain trick/method of stages?
7. How many non-negative integer solutions exist of the equation $x_1 + x_2 + \dots + x_m = n$?
8. What is the difference between a frequentist confidence interval and a Bayesian credible interval?
9. What is a one-sample Z test? How does it work? What is the difference between a one-sided and two-sided test?
10. Explain when a chi-squared test should be used. How does it work?
11. Define the posterior, prior and likelihood distributions. What is a conjugate prior?
12. What is the difference between a central 95% credible interval and a highest density credible interval? Explain how to write an algorithm to calculate these different types of credible intervals given a posterior. What if the posterior is not unimodal?
13. What is boxcar smoothing?
14. How can you calculate the coefficients of the best fitting polynomial when doing polynomial regression?
15. Write down the first order autoregressive model and calculate its general solution, assuming that it has been running for a long time and is in steady state. What condition is required on ϕ ?
16. Given an observed path, how can you calculate the joint likelihood function of a, b for the autoregressive model $x_k = ax_{k-1} + bs_k$, where the noise terms $\{s_k\}$ are independent Gaussians with zero mean and unit variance?
17. Comment on when it is appropriate to calculate a likelihood/posterior using a brute force approach, using MCMC, and using ABC. Given a complete observation of a stochastic SIR epidemic, what is the likelihood function of (β, μ) ?
18. Explain the basic MCMC algorithm outlined in lectures. What is a burn-in period?
19. Explain the ABC rejection sampling algorithm.

- 20.** What are the three main categories of ML algorithm explained in lectures? What are the differences between them?
- 21.** For binary data with a single input variable, how can logistic regression be used as a classification algorithm? What is the classification boundary?
- 22.** How can logistic regression be used to classify binary data with two input variables?
- 23.** Explain how K-nearest neighbour classification works. What are the potential issues with this approach?
- 24.** Explain the K-means clustering algorithm. Is this an example of a hierarchical or partition clustering algorithm?
- 25.** Explain how the output from a neuron is calculated from its inputs. Give an overview of how a neural network can be trained for a dataset such as the “grapefruit” example in lectures.