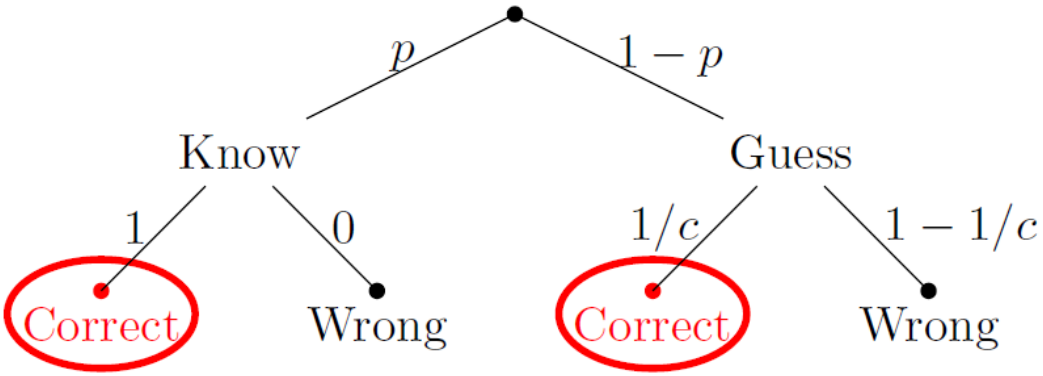
(200 pts total)

Section 1 – Short answer

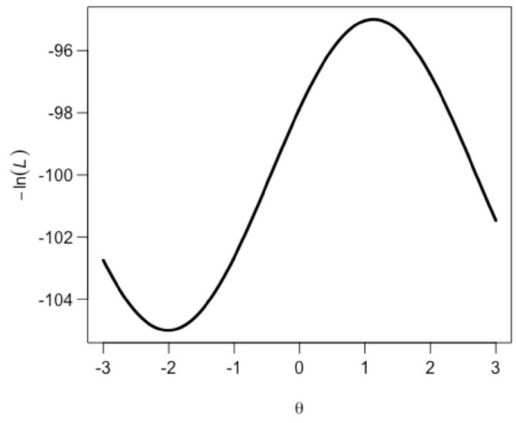
1. (5 pts) You are dealt one card from an unknown deck of cards, and that card is an Ace. Which hypothesis has a higher likelihood, H1: The deck is comprised entirely of Aces, or H2: The deck is a normal standard deck of cards? Why?

2. (15 pts) Imagine you are taking an exam. (I know that's tough, but just imagine…) The probability tree describing the probability of getting the correct answer given two scenarios (you know the answer, or you don't and you are guessing) is as follows.

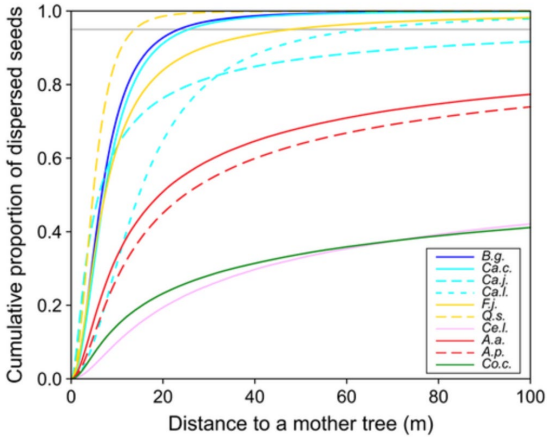


What is the probability that you know the answer conditional on having gotten it correct (i.e. $P(K|C)$)? (Hint: Use Bayes Theorem)

3. (5 pts) Based on the figure below, what is the (approximate) MLE for the parameter θ.



4. (5 pts) Plotted below is the cumulative probability density function associated with the dispersal of seeds from a "mother" tree (from Masaki et al. 2019). (The different curves represent different species.) Name one possible distribution that would fit these data.



5. (5 pts) What is the probability of making a Type I error when making k comparisons, assuming α_c is your cut-off for significance?

6. (15 pts) Suppose a variable Z takes values between 0 and 1 and has probability density function 2z. Compute Var[Z].

7. (10 pts) Complete the following limits:

$$\lim_{n\to\infty} Binom(n,p) \rightarrow$$

$$\lim_{\lambda\to\infty} Pois(\lambda) \rightarrow$$

8. (4 pts each) In an F-test to compare the variances of two samples, what effect does reducing the value of the significance level from 0.05 to 0.01 have on….

a. …the probability of committing a Type II error (i.e., $P(Fail\ to\ reject\ H_0|H_A)$)?

b. …the standard error of the variance for the first sample?

c. …the power of the test?

d. …the sample size?

e. …the confidence interval for the ratio of the variances?

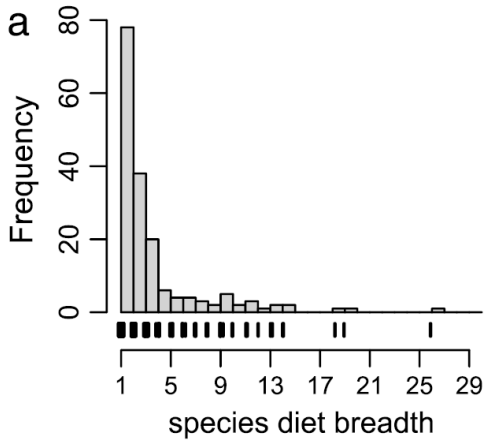
9. (10 pts) Fill in the blanks to construct the 1-α confidence interval for the mean of X where X~N(μ,σ^2) when the parametric variance is unknown.

$$P(\qquad\qquad \leq \mu \leq \qquad\qquad) = 1 - \alpha$$

10. (10 pts) What does it mean when we say that a certain approach to calculating confidence intervals has "good coverage"?

Section 2 – Long answer

5. (40 pts total) The Pareto distribution is a long-tailed distribution that has many uses in ecology, such as for modelling diet breadth among herbivores (i.e. the number of host plants eaten by a single species of herbivore)



From Kozubowski et al. (2015)

Let's assume we have data on diet breadth and assume the following Pareto distribution probability density function

$$P(X = x_i|\alpha) = \frac{\alpha x_i^{\alpha}}{S_o^{\alpha+1}}, x_i > S_o$$

where x_i is the number of plant species consumed in sample i, S_o is a constant, and α is a parameter of the distribution.

a) (5 pts) What is the support of this distribution?

b) (15 pts) Find the likelihood function for this model, given observations $X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n$.

c) (20 pts) Find the maximum likelihood estimate for $\hat{\alpha}$, given observations $X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n$.

5. (60 pts total) In an experiment to see if moths can retain memories through metamorphosis, 5th instar *Manduca sexta* caterpillars were trained to associate a mild electrical shock with a specific odor (ethyl acetate; EA). They were then tested to see whether they preferred clean air or EA-doped air as larvae and again as adult moths, after metamorphosis (Blackiston et al. 2008. Retention of memory through metamorphosis: can a moth remember what it learned as a caterpillar? PLoS ONE 3: e1736). The results of the experiment were as follows:

	Caterpillar treatment	
Adult response	Received shock treatment	Control
Chose clean air	32	25
Chose EA-doped air	9	21
Total	41	46

a) What is the best distribution to use for modelling the number of adults that chose clean air in each treatment? (5 pts)

For our analysis, we'll use the odds ratio to measure association between the caterpillar treatment and adult response. The odds O are defined as the ratio between the probability that an event happens and the probability that an event does not happen, i.e.

$$O = \frac{p}{1-p}$$

In this context, p is the probability of choosing clean air (this is arbitrary, but let's define it this way for consistency). The odds *ratio* (OR) compares the odds of choosing clean air under the two treatments, i.e.

$$OR = \frac{O_1}{O_2}$$

b) What is the odds-ratio (OR) for choosing clean air in this experiment? (5 pts)

c) Describe in words or in R code how you could use the bootstrap to obtain the standard error and a 95^{th} percentile confidence interval for the true (population) odds-ratio (OR). (10 pts).

d) What is one parametric test you could use to test the hypothesis that there is no difference (in the % that choose clean air) between the two treatments? (Include the test statistic and its distribution under the null hypothesis) (20 pts)

e) What is one non-parametric method of testing the hypothesis that there is no difference (in the % that choose clean air) between the two treatments? (Include the test statistic and how you would derive its distribution under the null hypothesis) (20 pts)