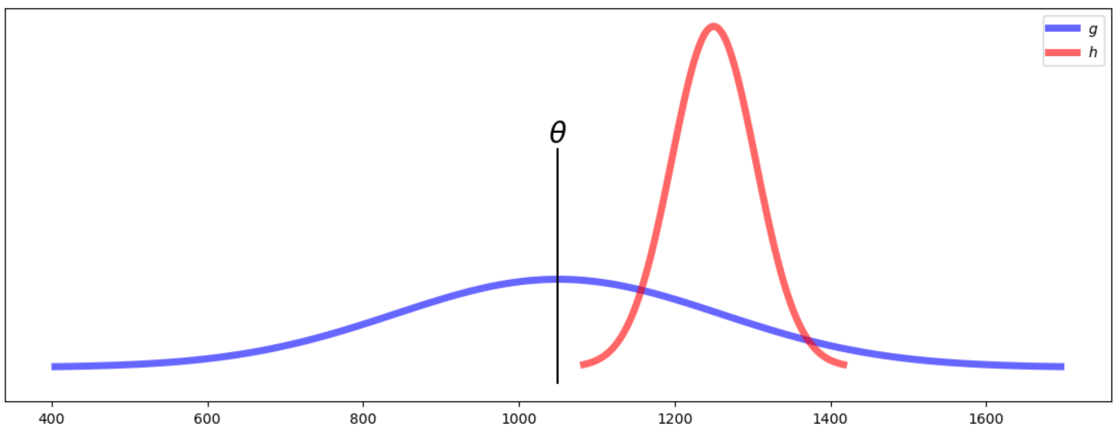
DS 122 Midterm 2

Name:

BUID:

Problem A (4 points)

Consider two estimators, $g()$ and $h()$, that both use data sampled from a parametric distribution to estimate that distribution's parameter θ. The plot below shows the distribution of values one obtains for both $g()$ and $h()$ over different data samples, where all samples are the same size (eg, each sample contains the same number n of data items).



Part 1

(2 points) Contrast the two estimators: which has higher bias, and which has higher variance?

Solution:

g has higher variance, and h has higher bias.

Part 2

(2 points) State which estimator(s) if any appear to be unbiased and why.

Solution:

g appears to be unbiased because its mean looks to be close to θ.

Problem B (10 points)

Assuming that we have a standard Binomial distribution of n independent Bernoulli trials each with a probability of success p, and let x_i be the outcome of the i-th trial, where $x_i = 1$ if the trial is successful and $x_i = 0$ otherwise, find the maximum likelihood estimator for the probability of success for this Binomial distribution.

Hint: start with probability of a single Bernoulli trial: $P(x, p) = p^x(1-p)^{1-x}$ and consider it's likelihood over n trials.

Solution:

Assuming that we have observed n independent Bernoulli trials with a probability of success p, and let x_i be the outcome of the i-th trial, where $x_i = 1$ if the trial is successful and $x_i = 0$ otherwise, then the likelihood function is given by:

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

Taking the logarithm of the likelihood function, we have:

$$\log L(p) = \sum_{i=1}^{n}[x_i \log(p) + (1-x_i)\log(1-p)]$$

To find the MLE for p, we differentiate the log-likelihood with respect to p and set it equal to zero:

$$\frac{\partial}{\partial p}\log L(p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\sum_{i=1}^{n}(1-x_i)}{1-p} = 0$$

Simplifying the above equation, we get:

$$\hat{p}\text{MLE} = \frac{\sum i=1^n x_i}{n}$$

Therefore, the MLE for the probability of success in a binomial distribution is the sample mean of the observed outcomes.

Problem C (10 points)

Consider the situation of parents who are expecting twins. So-called "identical" twins are twins with identical DNA (meaning they will be born the same sex), while "fraternal" twins are twins whose similarity at the DNA level is like any pair of siblings. The parents are curious about whether their twins are identical (but the babies are not born yet). They know that among the general population, about $1/3$ of twins are identical.

Part 1

(2 points) Define the two possible hypotheses corresponding to the two cases of interest to the parents.

Solution:

The twins are identical or the twins are not identical.

Part 2

(2 points) What is the prior probability of each hypothesis?

Solution:

$P(\text{Identical}) = 2/3$ and $P(\text{Not Identical}) = 1/3$

Part 3

(2 points) The parents then get an ultrasound and find out that the twins are the same birth sex. What is the likelihood of this outcome under each hypothesis?

Solution:

$P(D \mid \text{Identical}) = 1$ and $P(D \mid \text{Not Identical}) = 1/2.$

Part 4

(4 points) What is the posterior probability of each hypothesis? (That is, given knowledge of the ultrasound, what should the parents conclude)?

Solution:

$$P(\text{Identical} \mid D) = \frac{P(D \mid \text{Identical})P(\text{Identical})}{P(D \mid \text{Identical})P(\text{Identical}) + P(D \mid \text{Not Identical})P(\text{Not Identical})} = \frac{1/2 \cdot 2/3}{(1/2 \cdot 2/3) + (1 \cdot 1/3)} = \frac{1}{2}$$

And by a similar calculation, $P(\text{Not Identical} \mid D) = \frac{1}{2}$

Problem D (10 points)

You are given three dice: a six-sided die, a ten-sided die, and a twenty-sided die. You choose one of the dice at random, roll it, and observe that you get a nine.

Denote H_0 the hypothesis that you rolled the six-sided die, H_1 the hypothesis that you rolled the ten-sided die, and H_2 the hypothesis that you rolled the twenty-sided die. Fill out the following table.

hyp	P(H)	P(D|H)	P(H)P(D|H)	P(H|D)
H_0				
H_1				
H_2				

Solution:

hyp	P(H)	P(D|H)	P(H)P(D|H)	P(H|D)
H_0	1/3	0	0	0
H_1	1/3	1/10	1/30	2/3
H_2	1/3	1/20	1/60	1/3

Problem E (10 points)

You are playing Dungeons and Dragons. You are attacking 2d4 (meaning your attack is the sum of two four-sided dice), but you only hit your opponent if you exceed their defense roll which is 1d6 (meaning a single roll of a six-sided dice). You want to determine the probability you will hit your opponent. It turns out this is the probabilty of superiority of your dice roll over theirs. Breaking this down into two parts:

Part 1

(5 points) What is the probability distribution of the sum of two four-sided dice

Solution:

Every exact pair of rolls has a 1/16 chance of occuring (e.g. the probability of rolling a 3 on the first die and 4 on the second is 1/4 * 1/4 = 1/16). So we need to count how many ways each outcome can occur:

Outcome	Number of Ways	Probability
1	0	0
2	1 way (1,1)	1/16
3	2 ways (1,2 or 2,1)	2/16 = 1/8
4	3 ways (1,3 or 2,2 or 3,1)	3/16
5	4 ways (1,4 or 2,3 or 3,2 or 4,1)	4/16 = 1/4
6	3 ways (2,4 or 3,3 or 4,2)	3/16
7	2 ways (3,4 or 4,3)	2/16 = 1/8
8	1 ways (4,4)	1/16

Part 2

(5 points) What is the probability of superiority of of the sum of two four-sided dice over a single 6 sided dice?

Solution:

The probability of superiority is the sum of the probabilities of pairs of outcomes where the sum of two four-sided dice is greater than the 6 sided die. One option would be to list out all $7 * 6 = 42$ pairs of outcomes and sum their probabilities. We can do this slightly faster as follows:

- An outcome of 8 is always superior, so we can add 1/16 to our probability of superiority.
- Same for 7 so we can add 1/8
- 6 is superior to all rolls other than 6, so we can add $(3/16) * (5/6)$
- 5 is superior to all rolls other than 5 or 6, so we can add $(1/4) * (4/6)$
- And so on

This give us:

Probability of superiority = $1/16 + 1/8 + (3/16) * (5/6) + (1/4) * (4/6) +$
$(3/16) * (3/6) + (1/8) * (2/6) + (1/16) * (1/6) = 21/32$

Problem F (10 points)

You are running a (very small) experiment on coin flipping comparing a "control" coin to a "treated" coin. You flip each coin once and observe the first coin comes up heads and the second coin comes up tails. Let's call the probability the "control" coin comes up heads p_1 and the probability the treated coin comes up heads p_2.

You want to estimate the parameters p_1 and p_2 but you are short on time so you only consider the possibility that each coin is either fair ($p = 0.5$) rigged for tails ($p = 0.1$) or rigged for heads ($p = 0.9$). Starting with a uniform prior, what are the posterior probabilities for the two parameters for the three possible hypotheses?

Solution:

"Control" coin which came up heads:

Hypothesis	Prior	Likelihood	Unnormalized Posterior	Posterior
Coin is rigged for tails, $p_1 = 0.1$	1/3	0.1	1/30	1/15
Coin is fair, $p_1 = 0.5$	1/3	0.5	1/6	1/3
Coin is rigged for heads, $p_1 = 0.9$	1/3	0.9	9/30=3/10	3/5

"Treated" coin which came out tails:

Hypothesis	Prior	Likelihood	Unnormalized Posterior	Posterior
Coin is rigged for tails, $p_2 = 0.1$	1/3	0.9	9/30=3/10	3/5
Coin is fair, $p_2 = 0.5$	1/3	0.5	1/6	1/3
Coin is rigged for heads, $p_2 = 0.9$	1/3	0.1	1/30	1/15