DS 122 Practice Midterm 1

Name:

BUID:

Problem A (10 points)

Part 1

(5 points) A box contains three coins: two regular coins and one fake two-headed coin ($P(H) = 1$). You pick a coin at random and toss it. What is the probability that it lands heads up?

Part 2

(5 points) You pick a coin at random and toss it, and get heads. What is the probability that it is the two-headed coin?

Problem B (20 points)

Let X be a discrete random variable having the following distribution:

$$p_X(x) = \begin{cases} 1/3 \text{ when } x = 1, 2, 3 \\ 0 \text{ otherwise} \end{cases}$$

Part 1

(5 points) Find $E[X]$.

Part 2

(5 points) Find $Var(X)$.

Part 3

(10 points) Define the random variable Y to be $3X^2 - X$. Find the expected value of Y.

Problem C (10 points)

Show that for two independant random variables X and Y

$$Var(X - Y) = Var(X) + Var(Y)$$

Hint: Start with the fact that $Var(X) = E[X^2] - E[X]^2$

Problem D (10 points)

You challenge a friend to a free throw shooting contest. For each round of this contest you each take a shot. If you both make it or neither of you make it, no one wins. If only only you score, your friend gives you a dollar, and if only they score, you give them a dollar. You know that you're a better shooter than your friend, you make 80% of your shots while they only make 70% of their shots. Assume all shots are i.i.d.

Part 1

(5 points) How many rounds do you need to play before you expect to make 100 dollars?

Part 2

(5 points) After 100 rounds, there is a 68% chance that your earnigns will be in what range?

Problem E (5 points)

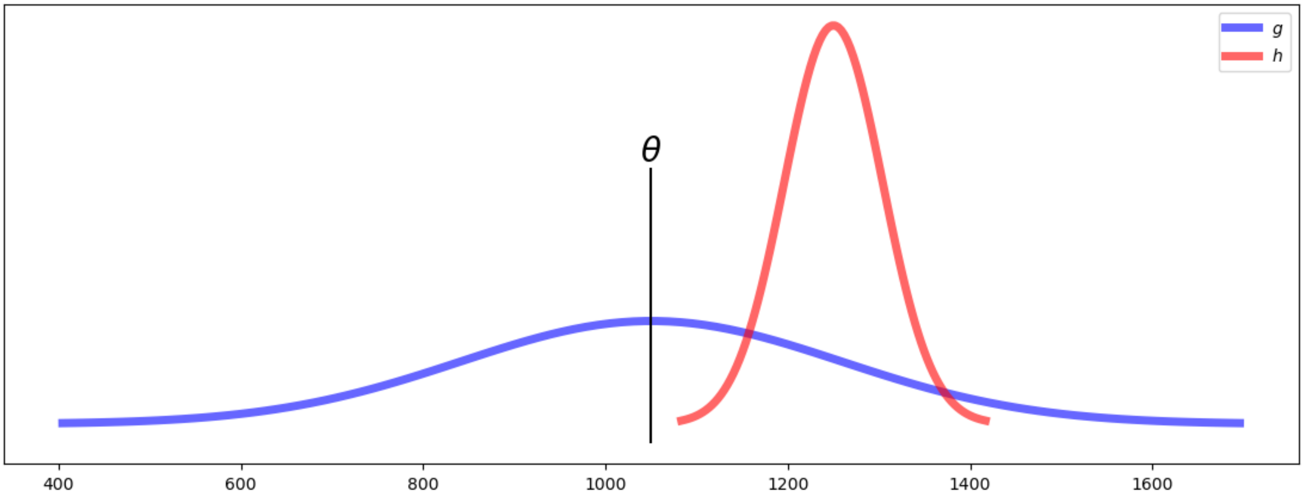
You sample heights of college freshman and college sophomores at random. You manage to collect 45 freshman heights and 61 sophomore heights. You immediately take your measuremnts and calculate their means and standard deviations. You find that the mean of the first sample is 66 inches with a standard deviation of 6 and the mean of the second sample is 70 inches with a standard deviation of 5. Note: assume the distribution of heights in the overall population is normal.

(4 points) What are the 95% confidence intervals for the average heights of the two populations?

(1 point) In words, what you can conclude about the mean heights of the populations relative to each other based on your confidence intervals?

Problem F (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?

Part 2

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

Problem G (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.

Extra Practice Problems

Problem A

Suppose a factory produces light bulbs, and the lifespan of these light bulbs follows a distribution with a mean of 1,000 hours and a standard deviation of 100 hours.

Part 1

If you were to take a sample of 30 light bulbs, describe the distribution of the sample mean.

Part 2

Calculate the probability that the sample mean lifespan of these 30 light bulbs is more than 1,020 hours.

Part 3

How large does the sample size need to be if you want the standard error to be less than 5 hours?

Problem B

Suppose a random variable X can take on values {1,2,3} with probabilities {0.3,0.4,0.3}, respectively. Let Y = X^2

Part 1

List the possible values of Y and their respective probabilities.

Part 2

Calculate the expected value E[Y]

Part 3

Find the variance Var[Y]

Problem C

Suppose that the lifetime of a certain type of light bulbs is modeled by an exponential distribution with an unknown parameter λ. Five bulbs were tested and the following lifetimes were found: 2, 3, 1, 3, and 4 years.

Hint: The probability density function of an exponential distribution is given by

$$p(x) = \lambda e^{-\lambda x}.$$

Part 1

Assuming that the lifetimes observations are independent, find the likelihood function.

Part 2

Find the corresponding log-likelihood function.

Part 3

Find the maximum likelihood estimate of λ.

Problem D

Suppose $\{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ is an independent and identically distributed sample from the exponential distribution with rate λ. Consider the following estimator for the population mean, $\frac{1}{\lambda}$:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}.$$

Is $\hat{\theta}$ a biased estimator? Support your answer by explicitly by computing the bias of $\hat{\theta}$.