Homework Assignment 1b

Due: Friday, Feb. 2, 2024, 11:59 p.m. Mountain time

Total marks: 65

Question 1. [15 MARKS]

Let X be a random variable with outcome space $\Omega = \{a, b, c\}$ and $p(a) = 0.1, p(b) = 0.2$, and $p(c) = 0.7$. Let

$$f(x) = \begin{cases} 10 & \text{if } x = a \\ 5 & \text{if } x = b \\ 10/7 & \text{if } x = c \end{cases}$$

(a) [4 MARKS] What is $E[f(X)]$?

(b) [3 MARKS] What is $E[1/p(X)]$? (Note: This is an expectation about the probability)

(c) [4 MARKS] With the same outcome space but now for an arbitrary pmf p, what is $E[1/p(X)]$? (Note: This is an expectation about the probability)

(d) [4 MARKS] What is $E[f(X)^2]$ and $E[f(X)]^2$?

Question 2. [15 MARKS]

Suppose you have three coins. Coin A has a probability of heads of 0.75, Coin B has a probability of heads of 0.5, and Coin C has a probability of heads of 0.25.

(a) [5 MARKS] Suppose you flip all three coins at once, and let X be the number of heads you see (which will be between 0 and 3). What is the expected value of X, $E[X]$?

(b) [10 MARKS] Suppose instead you put all three coins in your pocket, select one at random, and then flip that coin 5 times. You notice that 3 of the 5 flips result in heads while the other 2 are tails (Note: you do not know in what sequence the 3 heads and 2 tails occured). What is the probability that you chose Coin C? Hint: Define random variable D as the observed data, and notice that you can compute $p(D = 3 \text{ heads and } 2 \text{ tails} | \text{Coin} = C)$. You can use a binomial distribution: $P(k \text{ heads and } n - k \text{ tails}) = \binom{n}{k} p^k (1 - p)^{n-k}$, where n is the number of coin flips and $p \in [0, 1]$ is the probability of heads.

Question 3. [10 MARKS]

Alberta Hospital occasionally has electrical problems. It can take some time to find the problem, though it is always found in no more than 10 hours. The amount of time is variable; for example, one time it might take 0.3 hours, and another time it might take 5.7 hours. The time (in hours) necessary to find and fix an electrical problem at Alberta Hospital is a random variable, say X, whose density is given by the following uniform distribution

$$p(x) = \begin{cases} \frac{1}{10} & \text{if } 0 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

Such electrical problems can be costly for the Hospital, more so the longer it takes to fix it. The cost of an electrical breakdown of duration x is x^3. What is the expected cost of an electrical breakdown? Show your work.

Question 4. [25 MARKS]

We have talked about the fact that the sample mean estimator $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimator of the mean μ for identically distributed $X_1, X_2, \ldots, X_n$: $\mathbb{E}[\bar{X}] = \mu$. The straightforward variance estimator, on the other hand, is not an unbiased estimate of the true variance σ^2: for $\bar{V}_b = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, we get that $\mathbb{E}[\bar{V}_b] = (1 - \frac{1}{n})\sigma^2$. Instead, the following bias-corrected sample variance estimator is unbiased: $\bar{V} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. This unbiased estimator is typically what is called the *sample variance*.

(a) [15 MARKS] Use the fact that $\mathbb{E}[\bar{V}] = \sigma^2$ to show that $\mathbb{E}[\bar{V}_b] = (1 - \frac{1}{n})\sigma^2$. **Hint:** The proof is short, it can be done in a few lines. To start, consider if you can rewrite $\bar{V}_b$ in terms of $\bar{V}$ and then use your expectation rules.

(b) [10 MARKS] We also discussed the variance of the sample mean estimator, and concluded that $\mathrm{Var}[\bar{X}] = \frac{1}{n}\sigma^2$, for iid variables with variance σ^2. We can similarly ask what the variance is of the sample variance estimator. Deriving the formula is a bit more complex for general random variables, so let's assume the X_i are zero-mean Gaussian. For zero-mean Gaussian X_i, we can use $\bar{V} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$, which is unbiased, i.e., $\mathbb{E}[\bar{V}] = \sigma^2$. (Note that this $\bar{V}$ is different from (a). Here the estimator subtracts the **true mean**, which we know is zero, from each sample. The previous estimator subtracted the **sample mean**. This is why this estimator is unbiased although it divides the sum by n while the previous estimator had to divide the sum by $n-1$ to be unbiased). Then we know that the following is true (though we omit the derivation): $\mathrm{Var}[\bar{V}] = \frac{2(n-1)}{n^2}\sigma^4$.

This variance enables us to use Chebyshev's inequality, to get a confidence interval. Recall that Chebyshev's inequality states that for a random variable Y with known variance v, we know that

$$\Pr(|Y - \mathbb{E}[Y]| < \epsilon) > 1 - \frac{v}{\epsilon^2}$$

For some confidence level δ, this means we can set $\delta = v/\epsilon^2$ and solve for ϵ in terms of the two knowns: δ and v. Derive the confidence interval for $\bar{V}$, using Chebyshev's inequality.

Homework policies:

Your assignment should be submitted as one pdf document on eClass. The pdf must be written legibly and scanned or must be typed (e.g., Latex). This .pdf should be named Firstname_LastName_Sol.pdf,

Because assignments are more for learning, and less for evaluation, grading will be based on coarse bins. **The grading is atypical**. For grades between (1) 81-100, we round-up to 100; (2) 61-80, we round-up to 80; (3) 41-60, we round-up to 60; and (4) **0-40, we round down to 0**. The last bin is to discourage quickly throwing together some answers to get some marks. The goal for the assignments is to help you learn the material, and completing less than 50% of the assignment is ineffective for learning.

We will not accept late assignments. There is no late penalty policy. The assignments must be submitted electronically via eClass on time, by 11:59 pm Mountain time on the due date. There is a grace period of 48 hours when assignments will be accepted. No submissions will be accepted after 48 hours after the deadline, and the assignment will be considered as incomplete if not submitted.

All assignments are individual. All the sources used for the problem solution must be acknowledged, e.g. web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously; for detailed information see the University of Alberta Code of Student Behaviour.

Good luck!