CS1450 Final Review Problems
Fall 2018

Problems 1-6 are based on the midterm review. Identical problems are marked [recap]. Please consult previous recitations and textbook problems for additional practice.

1. Independence: Suppose $\mathcal{X}$, $\mathcal{Y}$ are independent.

   (a) [recap] What is $P(\mathcal{X} = x \land \mathcal{Y} = y)$?

   (b) [recap] What is $P(\mathcal{X} = x \lor \mathcal{Y} = y)$?

   (c) [recap] Prove that $P(\mathcal{X} = x) + P(\mathcal{Y} = y) - P(\mathcal{X} = x \land \mathcal{Y} = y) = 1 - P(\mathcal{X} \neq x \land \mathcal{Y} \neq y)$. How else could we describe this quantity?

   (d) Now suppose $\mathcal{Z}$ and $\mathcal{W}$ are both uniformly distributed (independently) on $\{0,1\}$, and take $\mathcal{Z} = \mathcal{X} \oplus \mathcal{Y}$, where $\oplus$ denotes exclusive or (in other words $\mathcal{Z} = 0$ if $\mathcal{X} = \mathcal{Y}$, and $\mathcal{Z} = 1$ otherwise). Is each pair of $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ independent? Are $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ mutually independent?

2. Variance and Covariance: We will now derive some classic properties of variance and covariance. Assume real-valued random variables $\mathcal{X}$ and $\mathcal{Y}$.

   (a) [recap] If $\mathcal{X}$ and $\mathcal{Y}$ are independent, is it always the case that $\text{Cov}(\mathcal{X}, \mathcal{Y}) = 0$ (under the independence assumption)? Either prove the statement or demonstrate that it is false with a counterexample.

   Hint: recall that $E[\mathcal{X}\mathcal{Y}] = E[\mathcal{X}]E[\mathcal{Y}] \iff \mathcal{X}, \mathcal{Y}$ independent.

   (b) Prove the inequality $\forall(\alpha\mathcal{X} + \beta\mathcal{Y}) = \alpha^2 E[\mathcal{X}] + \beta^2 E[\mathcal{Y}] + 2\alpha\beta \text{Cov}(\mathcal{X}, \mathcal{Y})$, for any $\alpha, \beta \in \mathbb{R}$.

   (c) Prove the inequality $\text{Cov}(\mathcal{X} + \mathcal{Y}, \mathcal{Z}) = \text{Cov}(\mathcal{X}, \mathcal{Z}) + \text{Cov}(\mathcal{Y}, \mathcal{Z})$.

   (d) Derive an equation for $\text{Cov}(\mathcal{X}, \mathcal{Y})$ in terms of $E[\mathcal{X}^2], E[\mathcal{Y}], E[\mathcal{X}\mathcal{Y}]$.

   Hint: This result is similar to the well-known $\forall[\mathcal{X}] = E[\mathcal{X}^2] - (E[\mathcal{X}])^2$.

3. Problems with Dice: Suppose I roll a fair 4 sided die and a fair 8 sided die (independently), each marked with consecutively increasing integers, starting from 1. Let $\mathcal{X}$, $\mathcal{Y}$ be the random variables associated with the number rolled on the 4-sided and 8-sided die, respectively.

   (a) [recap] What are the expectations and variances of $\mathcal{X}$ and $\mathcal{Y}$?

   (b) [recap] What is the expectation and variance of $\mathcal{X} + \mathcal{Y}$?

   (c) What are the expectation and variance of $\min(\mathcal{X}, \mathcal{Y})$?

   (d) [recap] Suppose I flip a fair coin, if heads I roll the 4 sided die, and if tails, I roll the 8 sided die. What are the expectation and variance of the roll value?

   (e) [recap] What is the probability that the 8-sided die is at least 7 and the 4-sided die is even?

   (f) [recap] What is the probability that the 8-sided die is at least 7 or the 4-sided die is even?

   (g) [recap] Suppose I observe that $\mathcal{Y} \geq 7$, and the $\mathcal{X}$ is even. What is the probability that $\mathcal{X} + \mathcal{Y} = 9$?

4. Conditional Exponentials and Conditional Expectations: Suppose $\mathcal{X}$ is exponentially distributed with rate parameter $\lambda$.

   (a) [recap] Suppose we observe that $\mathcal{X} > \alpha$ for some $\alpha > 0$. Given this observation, how is $\mathcal{X}$ distributed?

   (b) [recap] What is $E[\mathcal{X} \mid \mathcal{X} > \alpha]$ for some $\alpha > 0$. Try to use the previous result to solve this problem.

   (c) [recap] Solve $\text{[14]}$ again, this time using only integration and calculus (Hint: use integration by parts). Do your answers agree? Which strategy do you prefer?

   (d) [recap] Suppose I flip unbiased coins until achieving the first heads. Then for each coin I flipped, I draw an exponential random variable with rate parameter $\lambda$ and sum them. What is the expectation of their sum?
5. Gaussians, Expectations and Variances: Suppose $\mathcal{X}$ and $\mathcal{Y}$ are jointly normally distributed with means and variances $\mu_{\mathcal{X}}, \sigma^2_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}, \sigma^2_{\mathcal{Y}}$, respectively, and let $\sigma_{\mathcal{X},\mathcal{Y}} = \text{Cov}(\mathcal{X}, \mathcal{Y})$.

**Hint:** Recall that $\mathcal{V}[\alpha \mathcal{X} + \beta \mathcal{Y}] = \alpha^2 \mathcal{V}[\mathcal{X}] + \beta^2 \mathcal{V}[\mathcal{Y}] + 2\alpha\beta \text{Cov}(\mathcal{X}, \mathcal{Y})$.

(a) [recap] What are the expectation and variance of $\mathcal{X} + \mathcal{Y}$? **Bonus:** How is this quantity distributed?

(b) [recap] What is the variance of $-\alpha \mathcal{Y}$? How is this quantity distributed?

(c) [recap] What are the expectation and standard deviation of $\alpha \mathcal{X} - \beta \mathcal{Y}$? **Bonus:** How is this quantity distributed?

(d) [recap] Suppose $\mathcal{X}, \mathcal{Y}$ are independent. What is the joint density $f_{\mathcal{X}, \mathcal{Y}}(x, y)$? Simplify your solution within reason.

6. Fun with Probability Density Functions: Suppose $\mathcal{X}$ is distributed with PDF

$$f_{\mathcal{X}}(x) = \frac{1}{Z} \begin{cases} x & : x \in [0, 1) \\ 1 & : x \in [1, 3) \\ 0 & : \text{otherwise} \end{cases},$$

with some normalization constant $Z$ such that $\int_{\mathbb{R}} f_{\mathcal{X}}(x) \, dx = \int_{-\infty}^{\infty} f_{\mathcal{X}}(x) = 1.$

(a) [recap] What is the value of $Z$? Try to compute this quantity using geometry and using calculus. Make sure your answers agree!

(b) [recap] What is the CDF of $\mathcal{X}$?

(c) [recap] Suppose $\mathcal{Y} = \lfloor \mathcal{X} \rfloor$, where $\lfloor \cdot \rfloor$ is the floor operator that takes any real number onto the largest integer not greater than its argument. What are the PMF and CMF of $\mathcal{Y}$?

**Hint:** Recall that for $\mathcal{Y} = g(\mathcal{X})$, it holds that $P(\mathcal{Y} = y) = P(\mathcal{X} \in \{x \mid g(x) = y\})$.

(d) [recap] Without appealing to computing their values, how do $E[\mathcal{X}]$ and $E[\mathcal{Y}]$ compare? Defend your answer.

(e) [recap] **Bonus:** Suppose I draw (sample) a random value $\mathcal{X}_i$ with density $f_{\mathcal{X}}$. For any $i > 0$, if $\mathcal{X}_i > 1$, I discard it and restart the process for a new $\mathcal{X}_{i+1}$ drawn with density $f_{\mathcal{X}}$ with probability $\max(0, \min(2-x, 1))$, otherwise I let $\mathcal{X}_i$ equal $\mathcal{X}_i$.

Prove that $f_{\mathcal{X}}(x) = f_{\mathcal{X}}(x) P(\mathcal{X}_1 \text{ is discarded}) + f_{\mathcal{X}}(x) P(\mathcal{X}_1 \text{ is kept } | \mathcal{X}_1)$.

(f) [recap] **Bonus:** Use the solution to part (e) to derive $f_{\mathcal{X}}(\cdot)$.

7. Linear Regression: Suppose data points $x \in \mathbb{R}^{n \times m}$ and labels $y \in \mathbb{R}^m$, where each $x_i$ is distributed uniformly on $[-1, 1]^2$, and each $y_i \sim \mathcal{N}(x_{i,1} + x_{i,2}, \sigma^2)$. Define the risk of linear regression over a sample $(x, y)$ of size $m$ as

$$\tilde{R}_m(x, y) = \min_{\alpha \in \mathbb{R}^2, \alpha_0 \in \mathbb{R}} \frac{1}{m} \sum_{i=1}^{m} (x_i \cdot w + \alpha_0 - y_i)^2.$$

(a) As a function of $m$, what is $E_{x,y} \left[ \frac{1}{m} \sum_{i=1}^{m} (x_i \cdot w - y_i)^2 \right]$ for $w = (1, 1)$?

(b) As a function of $m$, what is $E_{x,y} \left[ \min_{\alpha_0 \in \mathbb{R}} \frac{1}{m} \sum_{i=1}^{m} (x_i \cdot w + \alpha_0 - y_i)^2 \right]$ for $w = (1, 1)$?

**Hint:** This formula resembles the sample variance formula. Can you relate $\alpha_0$ to the sample mean?

(c) What is $E_{x,y} \left[ \tilde{R}_m(x, y) \right]$, for $m \in \{1, 2, 3\}$? Does this trend continue as $m$ increases? What about for higher-dimensional data?

(d) Interpret the meaning of and derive the value of

$$\lim_{m \to \infty} \arg\min_{\alpha \in \mathbb{R}^2, \alpha_0 \in \mathbb{R}} \frac{1}{m} \sum_{i=1}^{m} (x_i \cdot w + \alpha_0 - y_i)^2.$$

(e) Interpret the meaning of and derive the value of

$$\lim_{m \to \infty} E_{x,y} \left[ \tilde{R}_m(x, y) \right].$$
(f) Prove or disprove the following statement:

\[
E_{x,y} [ \hat{R}_m(x, y)] \leq E_{x,y} \left[ \min_{w \in \mathbb{R}^2} \frac{1}{m} \sum_{i=1}^{m} (x_i \cdot w - y_i)^2 \right].
\]

What does this tell you about including the bias term \( w_0 \) in linear regression?

(g) Bonus: Prove or disprove the following statement:

\[
\forall m > 1 : E_{x,y} [ \hat{R}_{m-1}(x, y)] \leq E_{x,y} [ \hat{R}_m(x, y)].
\]

8. Asymptotic Average Theorems: The Pareto distribution with parameter \( \alpha \) has density \( \rho_{\alpha}(x) = \frac{\alpha}{x^{\alpha+1}} \) for \( x \geq 1 \), and 0 otherwise. Take \( \mathcal{X} \) to be Pareto distributed with \( \alpha = 1 \), \( \mathcal{Y} \) to be Pareto distributed with \( \alpha = 2 \), and \( \mathcal{Z} \) to be Pareto distributed with \( \alpha = 3 \).

**Hint:** Use 8a and 8b and asymptotic theorems to solve 8c and 8d. Pay attention to the assumptions made by each theorem! We recommend against evaluating these limits directly.

(a) Derive the expectation of \( \mathcal{X} \), \( \mathcal{Y} \), and \( \mathcal{Z} \).

(b) Derive the variance of \( \mathcal{X} \), \( \mathcal{Y} \), and \( \mathcal{Z} \).

(c) For \( x \) drawn i.i.d. from \( \mathcal{X} \), \( \mathcal{Y} \), and \( \mathcal{Z} \) (separately), what is \( \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} x_i \)? Prove your answer.

(d) For \( x \) drawn i.i.d. from \( \mathcal{X} \), \( \mathcal{Y} \), and \( \mathcal{Z} \) (separately), how is \( \lim_{m \to \infty} \frac{1}{\sqrt{m}} \sum_{i=1}^{m} x_i \) distributed? Prove your answer.

9. Markov Chains: Consider a Markov chain with states \( \{1, 2, 3\} \) and transition matrix

\[
T = \begin{bmatrix}
0.6 & 0.4 & 0 \\
0.2 & 0 & 0.8 \\
0 & 0.8 & 0.2
\end{bmatrix}.
\]

You are not required to fully simplify your answers for this problem.

(a) Find the stationary distribution of the Markov chain.

(b) Find the probability of being in state 3 after 3 steps if the chain begins at state 0.

(c) Find the probability distribution over states after 4 steps if the chain begins at a state chosen uniformly at random.

10. Hypothesis Testing: Suppose that under the null hypothesis, a server receives requests such that the time \( \mathcal{X} \) between them is exponentially distributed (independently) with rate parameter \( \lambda \).

(a) Develop a 1-tailed test based on a single request \( \mathcal{X} \) that falsely rejects the null hypothesis with probability \( \delta \) if an observed request interval is too small.

(b) Develop a 2-tailed test based on a single request that falsely rejects the null hypothesis with probability \( \delta \) if an observed request interval is too small or too large. Select your tails so they have equal probability under the null hypothesis.

(c) How do the 1-tailed tests for large and small time intervals and the 2-tailed test differ? When would you apply each type of test?

(d) Now suppose you wait for \( n \) requests, each occuring \( \mathcal{X}_i \) units of time after the previous. Take \( \min_{i=1}^{n} \mathcal{X}_i \) to be your new test statistic. Construct a 2-tailed test based on this statistic.

**Hint:** Recall that the rate of the minimum of \( n \) independent rate \( \lambda \) exponential random variables is itself exponential with rate \( n\lambda \).

(e) Again wait for \( n \) requests, and take \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{X}_i \) to be your new test statistic. Using the central limit theorem, how is this statistic distributed (asymptotically) under the null hypothesis? Now develop a 2-tailed asymptotic test using this statistic.

(f) Suppose that under \( H_0 \), \( \lambda = 2 \), and under \( H_1 \), \( \lambda = 3 \). As a function of \( n \) and \( \delta \), what is the probability of rejecting the null hypothesis given \( H_1 \) for your minimum and Gaussian sum tests?