Homework 0: Review Problems
CS/Stat 184

Harvard University
Due: September 8^{th} at 11:59 PM

1 Policies *[0 points]*

Please read these policies. **Please answer the three questions below and include your answers marked in a "problem 1" in your solution set.** Homeworks which do not include these answers will not be graded.

Gradescope submission: When submitting your HW report, please tag your pages correctly as is requested in gradescope. Untagged homeworks will not be graded, until the tagging is fixed.

Readings: Read the notes and required material.

Submission format: 1) Submit your report as a *single* pdf file on Gradescope. We highly encourage you to use LaTex. We will also accept neatly written handwritten homework. 2) Zip your code (python files, and a README.md explaining the functionality of each python file) for the programming component, and submit it on Gradescope.

Collaboration: It is acceptable for you to discuss problems with other students; it is not acceptable for students to look at another students written answers. Each student must understand, write, and hand in their own answers.

Acknowledgments: If students find out solutions in published material, on the web, or from other textbooks, this must be acknowledged. If students find proofs in existing papers, it is ok to use these for guidance; students must acknowledge this, and students should first make an attempt at the answer on their own. All students must understand all the written steps that they write.

1.1 List of Collaborators

List the names of all people you have collaborated with and for which question(s).

1.2 List of Acknowledgements

If you find an assignment's answer or use a another source for help, acknowledge for which question and provide an appropriate citation (there is no penalty, provided you include the acknowledgement). If not, then write "none".

1.3 Certify that you have read the instructions

Write "I have read these policies" to certify this.

2 Certify that you have read the website *[0 points]*

Please read the course policies on the website (up until the Lecture Notes section) and write"I have read the course policies on the website". It is your responsibility to understand and follow these policies. If further clarification is needed, please post to the discussion board.

3 Bayes Rule

1. ***[3 points]*** After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. What are the chances that you actually have the disease? (Show your calculations as well as giving the final result.)

4 Probability

1. ***[3 points]*** (independence and dependence) Let X and Y be real independent random variables with PDFs given by f and g, respectively. Let h be the PDF of the random variable $Z = X + Y$. Show that $h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$. (If you are more comfortable with discrete probabilities, you can instead derive an analogous expression for the discrete case, and then you should give a one sentence explanation as to why your expression is analogous to the continuous case.).
2. (conditional probabilities) Suppose X and Y are both independent and uniformly distributed on $[0, 1]$ (i.e. $f(x) = g(x) = 1$ for $x \in [0, 1]$ and 0 otherwise). Let h be the PDF of the random variable $Z = X + Y$. For these given explicit distributions,
   (a) ***[3 points]*** What is h?
   (b) ***[3 points]*** What is $\mathbb{P}\left(X \leq \frac{1}{2} \mid X + Y \geq \frac{5}{4}\right)$?
3. ***[3 points]*** (change of variable) A random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is Gaussian distributed with mean μ and variance σ^2. Given that for any $a, b \in \mathbb{R}$, we have that $Y = aX + b$ is also Gaussian, find a, b such that $Y \sim \mathcal{N}(0, 1)$.
4. For any two random variables X, Y the *covariance* is defined as $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. You may assume X and Y take on a discrete values if you find that is easier to work with.
   (a) ***[3 points]*** If $\mathbb{E}[Y|X = x] = x$ show that $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.
   (b) ***[3 points]*** If X, Y are independent show that $\text{Cov}(X, Y) = 0$.
5. If $f(x)$ is a PDF, the cumulative distribution function (CDF) is defined as $F(x) = \int_{-\infty}^{x} f(y)dy$. For any function $g : \mathbb{R} \to \mathbb{R}$ and random variable X with PDF $f(x)$, recall that the expected value of $g(X)$ is defined as $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(y)f(y)dy$. For a boolean event A, define $\mathbf{1}\{A\}$ as 1 if A is true, and 0 otherwise. Thus, $\mathbf{1}\{x \leq a\}$ is 1 whenever $x \leq a$ and 0 whenever $x > a$. Note that $F(x) = \mathbb{E}[\mathbf{1}\{X \leq x\}]$. Let $X_1, \ldots, X_n$ be *independent and identically distributed*

random variables with CDF $F(x)$. Define $\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \le x\}$. Note, for every x, that $\widehat{F}_n(x)$ is an *empirical estimate* of $F(x)$.

(a) ***[3 points]*** For any x, what is $\mathbb{E}[\widehat{F}_n(x)]$?

(b) ***[3 points]*** For any x, the variance of $\widehat{F}_1(x)$ is $\mathbb{E}[(\widehat{F}_1(x) - F(x))^2]$. Show that Variance$(\widehat{F}_1(x)) = F(x)(1 - F(x))$.

(c) ***[6 points]*** For any x, the variance of $\widehat{F}_n(x)$ is $\mathbb{E}[(\widehat{F}_n(x) - F(x))^2]$. Show that Variance$(\widehat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$.

(d) ***[6 points]*** Using your answer to c, show that for all $x \in \mathbb{R}$, we have $\mathbb{E}[(\widehat{F}_n(x) - F(x))^2] \le \frac{1}{4n}$.

5 Geometry and Linear Algebra

1. (Hyperplanes) Assume w is a n-dimensional vector and b is a scalar. A hyper-plane in $\mathbb{R}^n$ is the set $\{x : x \in \mathbb{R}^n, \text{ s.t. } w^\top x + b = 0\}$.

   (a) ***[3 points]*** ($n = 2$ example) Draw the hyperplane for $w = [-1,\ 2]^\top$, $b = 2$. Label your axes.

   (b) ***[3 points]*** ($n = 3$ example) Draw the hyperplane for $w = [1,\ 1,\ 1]^\top$, $b = 0$. Label your axes.

   (c) ***[6 points]*** (distance) Given some $x_0 \in \mathbb{R}^n$, find the *squared* distance to the hyperplane defined by $w^\top x + b = 0$. In other words, solve the following optimization problem:

$$\begin{aligned} &\min_x \ ||x_0 - x||^2 \\ &\text{s.t. } \ w^\top x + b = 0 \end{aligned}$$

   Remember, we want the squared distance, not the closest x.

2. ***[6 points]*** (Rank) Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$.

   (a) What is the rank of A? Why?

   (b) What is a (minimal size) basis for the column span?

3. ***[6 points]*** (Linear Equations and Matrix Multiplication) Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = [-2,\ -2,\ -4]^\top$, and $c = [1,\ 1,\ 1]^\top$.

   (a) What is Ac?

   (b) What is the solution to the linear system $Ax = b$?

4. ***[3 points]*** (Linear Algebra) For possibly non-symmetric $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ and $c \in \mathbb{R}$, define $f(x, y) = x^\top \boldsymbol{A} x + y^\top \boldsymbol{B} x + c$ for $x, y \in \mathbb{R}^n$. Explicitly write out the function $f(x, y)$ in terms of the components $\boldsymbol{A}_{i,j}$ and $\boldsymbol{B}_{i,j}$ using appropriate summations over the indices.

6 Programming

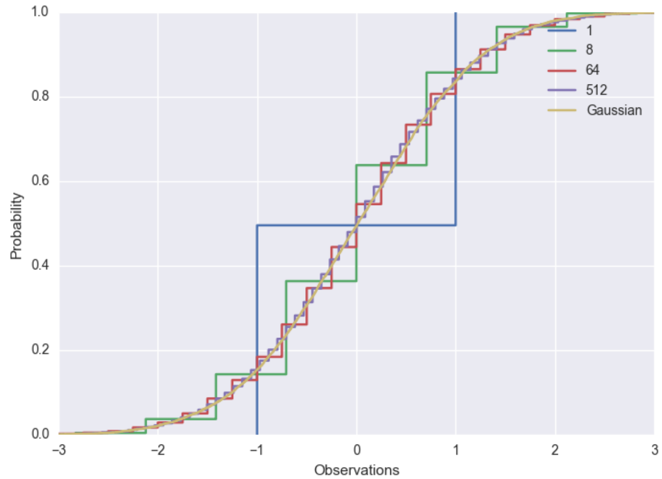
Please include your answer to the following questions in the report (the pdf file), and also submit your code on Gradescope as instructed. Your code should be well-commented. For this homework, feel free to name your python file(s), but remember to include a README.md that explains the functionality of each python file.

1. ***[6 points]*** For the values of A, b, and c as defined in Problem 5.3, use numpy to compute:
   (a) What is A^{-1}?
   (b) What is $A^{-1}b$? What is Ac? Take a screen shot of your answer.

2. ***[15 points]*** Two random variables X and Y have equal distributions if their CDFs, F_X and F_Y, respectively, are equal, i.e. for all x, $|F_X(x) - F_Y(x)| = 0$. The central limit theorem says that the sum of k independent, zero-mean, variance-$1/k$ random variables converges to a (standard) Normal distribution as k goes off to infinity. We will study this phenomenon empirically (you will use the Python packages Numpy and Matplotlib). Define $Y^{(k)} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} B_i$ where each B_i is equal to -1 and 1 with equal probability. It is easy to verify (you should) that $\frac{1}{\sqrt{k}} B_i$ is zero-mean and has variance $1/k$.
   (a) For $i = 1, \dots, n$ let $Z_i \sim \mathcal{N}(0, 1)$. If $F(x)$ is the true CDF from which each Z_i is drawn (i.e., Gaussian) and $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Z_i \leq x\}$, use the answer to problem 4.5 above to choose n large enough such that, for all $x \in \mathbb{R}$, $\sqrt{\mathbb{E}[(\widehat{F}_n(x) - F(x))^2]} \leq 0.0025$, and plot $\widehat{F}_n(x)$ from -3 to 3.
   (Hint: use `Z=numpy.random.randn(n)` to generate the random variables, and `import matplotlib.pyplot as plt;`
   `plt.step(sorted(Z), np.arange(1,n+1)/float(n))` to plot).
   (b) For each $k \in \{1, 8, 64, 512\}$ generate n independent copies $Y^{(k)}$ and plot their empirical CDF on the same plot as part a.
   (Hint: `np.sum(np.sign(np.random.randn(n, k))*np.sqrt(1./k), axis=1)` generates n of the $Y^{(k)}$ random variables.)

   Be sure to always label your axes.

Your plot should look something like the following (Tip: checkout `seaborn` for instantly better looking plots.)


1.0
0.8
0.6
0.4
0.2
0.0
Probability
−3
−2
−1
0
1
2
3
Observations
1
8
64
512
Gaussian