1. **(1 pts, 0.25 pts each)** *Linearity.* Are the following functions linear? Justify your answer.

   (a) $f(x) = \|x\|_2^2$

   (b) $f(x) = \|x\|_1$

   (c) $f(x) = \frac{1}{2}x^T Q x + p^T x + r$

   (d) $f(x) = c^T x + b^T A x$

2. **(1 pt, 0.25 each)** Using the properties of norms, verify that the following are norms, or prove that they are not norms

   (a) *Direct sum.* $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) = \sum_k x[k]$

   (b) *Sum of square roots, squared.* $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) = \left(\sum_{k=1}^{d} \sqrt{|x[k]|} \right)^2$

   (c) *(Shifted) entropy function* $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) = -\sum_{i=1}^{d} (x[i] + \frac{1}{2}) \log_2(x[i] + \frac{1}{2})$ where $-1/2 \le x[i] \le 1/2$ for all $i = 1, ..., d$

   (d) *Weighted 2-norm.* $f : \mathbb{R}^d \to \mathbb{R}$, $f(x) = \sqrt{\sum_{k=1}^{d} \frac{|x[k]|^2}{k}}$

3. **(1 pt, 0.25 each)** *Independent or not independent* Variables A and B are random variables for two distributions. Decide if A and B are independent. Justify your answer.

   (a) A and B are discrete random variables and have the following p.m.f.s

$$p_A(a) = \begin{cases} 0.25, & a = \text{red} \\ 0.25, & a = \text{blue} \\ 0.5, & a = \text{green} \end{cases}, \qquad p_B(b) = \begin{cases} 0.3, & b = \text{hat} \\ 0.3, & b = \text{T-shirt} \\ 0.2, & b = \text{skirt} \\ 0.2, & b = \text{shoes} \end{cases}$$

   and $p_{A,B}(a, b)$ are defined by the table below

	a = red	a = blue	a = green
b = hat	0.075	0.075	0.15
b = T-shirt	0.075	0.075	0.15
b = skirt	0.05	0.05	0.1
b = shoes	0.05	0.05	0.1

   (b) A and B are uniform distributions, where

$$f_A(a) = \begin{cases} 1 & -1 \le a \le 0 \\ 0 & \text{else,} \end{cases} \qquad f_B(b) = \begin{cases} 1 & 0 \le b \le 1 \\ 0 & \text{else,} \end{cases},$$

$$f_{A,B}(a, b) = \begin{cases} 4/3 & |a + b| \le 1/2, \ -1 \le a \le 0, \ 0 \le b \le 1 \\ 0 & \text{else,} \end{cases}$$

   (c) A follows the p.m.f.

$$p_A(a) = \begin{cases} 0.5, & a = 1 \\ 0.5, & a = -1 \end{cases}$$

   and $B = A \cdot C$ where

$$p_C(c) = \begin{cases} 0.9, & c = 1 \\ 0.1, & c = -1 \end{cases}$$

(d) A and B are Gaussian distributions, with the following properties:

$$\mathbb{E}[A] = 0, \quad \mathbb{E}[B] = 1, \quad \mathbb{E}[A^2] = 1, \quad \mathbb{E}[(B-1)^2] = 1/2, \quad \mathbb{E}[A(B-1)] = -1.$$

Writing in terms of the usual Gaussian distribution form, if we form a random vector as $X = \begin{bmatrix} A \\ B \end{bmatrix}$, then

$$\mu = \begin{bmatrix} \mathbb{E}[A] \\ \mathbb{E}[B] \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \mathbb{E}[(A - \mathbb{E}[A])^2] & \mathbb{E}[(A - \mathbb{E}[A])(B - \mathbb{E}[B])] \\ \mathbb{E}[(A - \mathbb{E}[A])(B - \mathbb{E}[B])] & \mathbb{E}[(B - \mathbb{E}[B])^2] \end{bmatrix}$$

4. **(2 pts)** *Probability and statistics.* I have 4 children, Alexa, Siri, Googs, and Zuckie. Every morning I tell them to put on their socks.

- Alexa only listens to me on Mondays and Thursdays and puts on her socks. The rest of the days, she puts on her socks only half of the time. She either puts on both her socks or none of her socks.
- Siri always runs and gets her socks, but only puts one sock on.
- Googs tells me all this random trivia about socks, but never puts on his socks.
- Zuckie wears both his socks 4/7 of the time and sells the rest of them to CambridgeAnalytica.

Assume the children all act independently. Round all answers to at least 3 significant digits.

(a) **(0.5 pts)** What are the chances that either Alexa or Zuckie is wearing a sock?

(b) **(0.5 pts)** On a random day, a girl is wearing a sock. What are the chances that it's Alexa?

(c) **(0.5 pts)** What is the expected number of socks being worn by each child?

(d) **(0.5 pts)** What is the variance in the number of socks being work by each child?

5. **(2 pts, 0.5 points each)** *Conditional independence vs independence.* Tom is a blue-gray cat with a bushy tail, and Jerry is a brown mouse with a rope-like tail. After many years of fighting, they both decided to settle down, and now have thriving families. Tom has 10 kids and Jerry has 40 kids. Tom's kids are all cats like him, with bushy tails. Half of Tom's kids are blue, while the other half is gray. Jerry's kids are all brown mice, with rope-like tails.

(a) I pick up a baby animal at random. What is the probability that ... (fill in the table)

fur \ tail	furry	rope-like
blue		
gray		
brown		

(b) Are the features "fur color" and "tail texture" independent or dependent, without knowing the type of animal? (Show mathematically.)

(c) Now Tom comes over and says, "I'm very proud of my baby girl, of whom you are holding." What is the probability that (fill in the table)

fur \ tail	furry	rope-like
blue		
gray		
brown		

(d) Are the features "fur color" and "tail texture" independent or dependent, now that I know the animal is Tom's cherished baby daughter? (Show mathematically.)

6. **(3 pts) Exponential distribution.** Wait time is often modeled as an exponential distribution, e.g.

$$\mathbf{Pr}(\text{I wait less than } x \text{ hours at the DMV}) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0, \end{cases}$$

and this cumulative density function is parametrized by some constant $\lambda > 0$. A random variable X distributed according to this CDF is denoted as $X \sim \exp[\lambda]$.

(a) **(0.25 pts)** In terms of λ, give the probability distribution function for the exponential distribution.

(b) **(0.25 pts)** Show that if $X \sim \exp(\lambda)$, then the mean of X is $1/\lambda$ and the variance is $1/\lambda^2$.

(You may use a symbolic integration tool such as Wolfram Alpha. If you do wish to do the integral by hand, my hint is to review integration by parts.)

(c) **(0.5 pts)** Now suppose I run a huge server farm, and I am monitoring the server's ability to respond to web requests. I have m observations of delay times, $x_1, ..., x_m$, which I assume are i.i.d., distributed according to $\exp[\lambda]$ for some λ. Given these m observations, what is the maximum likelihood estimate $\hat{\lambda}$ of λ?

(d) **(1 pt)** Given the estimate of $\hat{\lambda}$ in your previous question, is $1/\hat{\lambda}$ an unbiased estimate of the mean wait time? Is $1/\hat{\lambda}^2$ an unbiased estimate of the variance in wait time?

(e) **(1 pt)** Now let's consider $x_1, ..., x_m$ drawn i.i.d. from a *truncated* exponential distribution, e.g.

$$p_{\lambda,c}(x) = \begin{cases} 0 & \text{if } x > c \text{ or } x < 0 \\ \dfrac{\lambda \exp(-\lambda x)}{1 - \exp(-\lambda c)} & \text{else.} \end{cases}$$

Using Hoefdding's inequality, give a range of values that account for the uncertainty in your guess. That is, as a function of x_i, m and δ, give a range of values $[\hat{\lambda}_{\min}, \hat{\lambda}_{\max}]$ such that

$$\mathbf{Pr}(\hat{\lambda}_{\min} \leq \mathbb{E}[X] \leq \hat{\lambda}_{\max}) \geq 1 - \delta.$$

Challenge!

Generalizing the Cauchy Schwartz inequality

We saw in lecture the Cauchy Schwartz inequality, which says that for any two vectors $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$,

$$x^T y \leq \|x\|_2 \|y\|_2, \quad x^T y = \|x\|_2 \|y\|_2 \iff x = cy$$

for some positive scalar c.

1. Holder's inequality generalizes this to *dual norms*. That is, for any p and q where $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$x^T y \leq \|x\|_p \|y\|_q$$

where equality is reachable for a specific choice of x and y.

   (a) When $p = 1$, the corresponding q is $q = +\infty$, and $\|x\|_\infty$ is the max-norm, e.g. $\|x\|_\infty = \max_i |x_i|$. Prove Holder's inequality for this choice of p and q. That is, prove that for any x and any y,

$$x^T y \leq \|x\|_1 \|y\|_\infty.$$

   (b) For $p = 1$, $q = +\infty$, given x, list the entire set of possible y where $x^T y = \|x\|_1 \|y\|_\infty$. Hint: The rule for $x = [0, 0, ..., 0]^T$ is different than for x if all the values are nonzero.

2. We can further generalize Holder's inequality to include *conic constraints*. For example, suppose x is restricted to the nonnegative orthant, e.g. $x_i \geq 0$ for all i. Then

$$x^T y \leq \left(\sum_i x_i \right) \max_i y_i. \qquad (*)$$

   (a) Prove $(*)$.

   (b) List the set of y such that, given x, $(*)$ is true with equality.

3. The *singular value decomposition* of a matrix $X \in \mathbb{R}^{m \times n}$ decomposes X to its *singular values and vectors*. It is usually written as

$$X = \sum_{i=1}^r s_i u_i v_i^T$$

where $u_i \in \mathbb{R}^m$ are the *left singular values*, $v_i \in \mathbb{R}^n$ are the *right singular values*, and s_i are positive scalars. Here, r is the *rank* of X. Each singular vector are *orthonormal*, e.g.

$$u_i^T u_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else,} \end{cases} \qquad v_i^T v_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else.} \end{cases}$$

The *trace norm* of X is the sum of its singular values (denoted $\|X\|_*$). The *spectral norm* of X is its largest singular value (denoted $\|X\|_2$).

   (a) Prove the following generalization of the Cauchy Schwartz inequality for matrices $X \in \mathbb{R}^{m \times n}$ and $Y \in \mathbb{R}^{m \times n}$

$$\mathbf{tr}(X^T Y) \leq \|X\|_* \|Y\|_2.$$

   (b) Given X, describe the set of Y such that $\mathbf{tr}(X^T Y) = \|X\|_* \|Y\|_2$.