Please provide complete and well-written solutions to the following exercises.

Due January 20, 9AM, to be submitted in blackboard, under the Assignments tab.

Homework 1

Exercise 1. As needed, refresh your knowledge of proofs and logic by reading the following document by Michael Hutchings: http://math.berkeley.edu/∼hutching/teach/proofs.pdf

Exercise 2. As needed, take the following quizzes on logic and set theory:

http://scherk.pbworks.com/w/page/14864234/Quiz%3A%20Logic
http://scherk.pbworks.com/w/page/14864241/Quiz%3A%20Sets

(These quizzes are just for your own benefit; you don't need to record your answers anywhere.)

Exercise 3. Two people take turns throwing darts at a board. Person A goes first, and each of her throws has a probability of $1/4$ of hitting the bullseye. Person B goes next, and each of her throws has a probability of $1/3$ of hitting the bullseye. Then Person A goes, and so on. With what probability will Person A hit the bullseye before Person B does?

Exercise 4. Two people are flipping fair coins. Let n be a positive integer. Person I flips $n+1$ coins. Person II flips n coins. Show that the following event has probability $1/2$: Person I has more heads than Person II.

Exercise 5. Suppose a test for a disease is 99.9% accurate. That is, if you have the disease, the test will be positive with 99.9% probability. And if you do not have the disease, the test will be negative with 99.9% probability. Suppose also the disease is fairly rare, so that roughly 1 in $20,000$ people have the disease. If you test positive for the disease, with what probability do you actually have the disease?

Exercise 6 (Inclusion-Exclusion Formula)**.** In the Properties for Probability laws, we showed that $\mathbf{P}(A\cup B)=\mathbf{P}(A)+\mathbf{P}(B)-\mathbf{P}(A\cap B)$. The following equality is a generalization of this fact. Let Ω be a discrete sample space, and let $\mathbf{P}$ be a probability law on Ω. Prove the following. Let $A_1,\ldots,A_n\subseteq\Omega$. Then:

$$\mathbf{P}(\cup_{i=1}^n A_i)=\sum_{i=1}^n\mathbf{P}(A_i)-\sum_{1\leq i<j\leq n}\mathbf{P}(A_i\cap A_j)+\sum_{1\leq i<j<k\leq n}\mathbf{P}(A_i\cap A_j\cap A_k)$$
$$\cdots+(-1)^{n+1}\mathbf{P}(A_1\cap\cdots\cap A_n).$$

(Hint: begin with the identity $0=(1-1)^m=\sum_{k=0}^m(-1)^k\binom{m}{k}$, which follows from the Binomial Theorem. That is, $1=\sum_{k=1}^m(-1)^{k+1}\binom{m}{k}$. Now, let $x\in\Omega$ such that x is in exactly m of the sets $A_1,\ldots,A_n$. Compute the "number of times" that the element $x\in\Omega$ is counted for both sides of the Inclusion-Exclusion Formula.)

Exercise 7 (Stein Identity)**.** Let X be a standard Gaussian random variable, so that X has density $x \mapsto e^{-x^2/2}/\sqrt{2\pi}$, $\forall\, x \in \mathbf{R}$. Let $g\colon \mathbf{R} \to \mathbf{R}$ be a continuously differentiable function such that g and g' have polynomial volume growth. That is, $\exists\, a, b > 0$ such that $|g(x)|, |g'(x)| \leq a(1+|x|)^b$, $\forall\, x \in \mathbf{R}$. Prove the **Stein identity**

$$\mathbf{E}Xg(X) = \mathbf{E}g'(X).$$

Using this identity, recursively compute $\mathbf{E}X^k$ for any positive integer k.

Alternatively, for any $t > 0$, show that $\mathbf{E}e^{tX} = e^{t^2/2}$, i.e. compute the **moment generating function** of X. Then, using $\frac{d^k}{dt^k}|_{t=0}\mathbf{E}e^{tX} = \mathbf{E}X^k$ and using the power series expansion of the exponential, compute $\mathbf{E}X^k$ directly from the identity $\mathbf{E}e^{tX} = e^{t^2/2}$.

Exercise 8 (MAX-CUT)**.** The probabilistic method is a very useful way to prove the existence of something satisfying some properties. This method is based upon the following elementary statement: If $\alpha \in \mathbf{R}$ and if a random variable $X\colon \Omega \to \mathbf{R}$ satisfies $\mathbf{E}X \geq \alpha$, then there exists some $\omega \in \Omega$ such that $X(\omega) \geq \alpha$. We will demonstrate this principle in this exercise.

Let $G = (V, E)$ be an undirected graph on the vertices $V = \{1, \ldots, n\}$ so that the edge set E is a subset of unordered pairs $\{i, j\}$ such that $i, j \in V$ and $i \neq j$. Let $S \subseteq V$ and denote $S^c := V \setminus S$. We refer to (S, S^c) as a cut of the graph G. The goal of the MAX-CUT problem is to maximize the number of edges going between S and S^c over all cuts of the graph G.

Prove that there exists a cut (S, S^c) of the graph such that the number of edges going between S and S^c is at least $|E|/2$. (Hint: define a random $S \subseteq V$ such that, for every $i \in V$, $\mathbf{P}(i \in S) = 1/2$, and the events $1 \in S, 2 \in S, \ldots, n \in S$ are all independent. If $\{i, j\} \in E$, show that $\mathbf{P}(i \in S, j \notin S) = 1/4$. So, what is the expected number of edges $\{i, j\} \in E$ such that $i \in S$ and $j \notin S$?)

Exercise 9. Let $n \geq 2$ be a positive integer. Let $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$. For any $x, y \in \mathbf{R}^n$, define $\langle x, y\rangle := \sum_{i=1}^n x_i y_i$ and $||x|| := \langle x, x\rangle^{1/2}$. Let $S^{n-1} := \{x \in \mathbf{R}^n \colon ||x|| = 1\}$ be the sphere of radius 1 centered at the origin. Let $x \in S^{n-1}$ be fixed. Let v be a random vector that is uniformly distributed in S^{n-1}. Prove:

$$\mathbf{E}\,|\langle x, v\rangle| \geq \frac{1}{10\sqrt{n}}.$$

Exercise 10 (The Power Method)**.** This exercise gives an algorithm for finding the eigenvectors and eigenvalues of a symmetric matrix. In modern statistics, this is often a useful thing to do. The Power Method described below is not the best algorithm for this task, but it is perhaps the easiest to describe and analyze.

Let A be an $n \times n$ real symmetric matrix. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the (unknown) eigenvalues of A, and let $v_1, \ldots, v_n \in \mathbf{R}^n$ be the corresponding (unknown) eigenvectors of A such that $||v_i|| = 1$ and such that $Av_i = \lambda_i v_i$ for all $1 \leq i \leq n$.

Given A, our first goal is to find v_1 and λ_1. For simplicity, assume that $1/2 < \lambda_1 < 1$, and $0 \leq \lambda_n \leq \cdots \leq \lambda_2 < 1/4$. Suppose we have found a vector $v \in \mathbf{R}^n$ such that $||v|| = 1$ and

$|\langle v, v_1\rangle| > 1/n$. (From Exercise 9, a randomly chosen v satisfies this property.) Let k be a positive integer. Show that

$$A^k v$$

approximates v_1 well as k becomes large. More specifically, show that for all $k \geq 1$,

$$\left|\left|A^k v - \langle v, v_1\rangle \lambda_1^k v_1\right|\right|^2 \leq \frac{n-1}{16^k}.$$

(Hint: use the spectral theorem for symmetric matrices.)

Since $|\langle v, v_1\rangle|\, \lambda_1^k > 2^{-k}/n$, this inequality implies that $A^k v$ is approximately an eigenvector of A with eigenvalue λ_1. That is, by the triangle inequality,

$$\left|\left|A(A^k v) - \lambda_1(A^k v)\right|\right| \leq \left|\left|A^{k+1} v - \langle v, v_1\rangle \lambda_1^{k+1} v_1\right|\right| + \lambda_1 \left|\left|\langle v, v_1\rangle \lambda_1^k v_1 - A^k v\right|\right| \leq 2\frac{\sqrt{n-1}}{4^k}.$$

Moreover, by the reverse triangle inequality,

$$\left|\left|A^k v\right|\right| = \left|\left|A^k v - \langle v, v_1\rangle \lambda_1^k v_1 + \langle v, v_1\rangle \lambda_1^k v_1\right|\right| \geq \frac{1}{n} 2^{-k} - \frac{\sqrt{n-1}}{4^k}.$$

In conclusion, if we take k to be large (say $k > 10 \log n$), and if we define $z := A^k v$, then z is approximately an eigenvector of A, that is

$$\left|\left|A\frac{A^k v}{||A^k v||} - \lambda_1 \frac{A^k v}{||A^k v||}\right|\right| \leq 4n^{3/2} 2^{-k} \leq 4n^{-4}.$$

And to approximately find the first eigenvalue λ_1, we simply compute

$$\frac{z^T A z}{z^T z}.$$

That is, we have approximately found the first eigenvector and eigenvalue of A.

Remarks. To find the second eigenvector and eigenvalue, we can repeat the above procedure, where we start by choosing v such that $\langle v, v_1\rangle = 0$, $||v|| = 1$ and $|\langle v, v_2\rangle| > 1/(10\sqrt{n})$. To find the third eigenvector and eigenvalue, we can repeat the above procedure, where we start by choosing v such that $\langle v, v_1\rangle = \langle v, v_2\rangle = 0$, $||v|| = 1$ and $|\langle v, v_3\rangle| > 1/(10\sqrt{n})$. And so on.

Google's PageRank algorithm uses the power method to rank websites very rapidly. In particular, they let n be the number of websites on the internet (so that n is roughly 10^9). They then define an $n \times n$ matrix C where $C_{ij} = 1$ if there is a hyperlink between websites i and j, and $C_{ij} = 0$ otherwise. Then, they let B be an $n \times n$ matrix such that B_{ij} is 1 divided by the number of 1's in the i^{th} row of C, if $C_{ij} = 1$, and $B_{ij} = 0$ otherwise. Finally, they define

$$A = (.85)B + (.15)D/n$$

where D is an $n \times n$ matrix all of whose entries are 1.

The power method finds the eigenvector v_1 of A, and the size of the i^{th} entry of v_1 is proportional to the "rank" of website i.

Exercise 11. Let X_1, Y_1 be random variables with joint PDF f_{X_1,Y_1}. Let X_2, Y_2 be random variables with joint PDF f_{X_2,Y_2}. Let $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ and let $S\colon \mathbf{R}^2 \to \mathbf{R}^2$ so that $ST(x,y) = (x,y)$ and $TS(x,y) = (x,y)$ for every $(x,y) \in \mathbf{R}^2$. Let $J(x,y)$ denote the determinant of the Jacobian of S at (x,y). Assume that $(X_2, Y_2) = T(X_1, Y_1)$. Using the change of variables formula from multivariable calculus, show that

$$f_{X_2,Y_2}(x,y) = f_{X_1,Y_1}(S(x,y))\,|J(x,y)|\,.$$

Exercise 12. Suppose I tell you that the following list of 20 numbers is a random sample from a Gaussian random variable, but I don't tell the mean or standard deviation.

5.1715, 3.2925, 5.2172, 6.1302, 4.9889, 5.5347, 5.2269, 4.1966, 4.7939, 3.7127
5.3884, 3.3529, 3.4311, 3.6905, 1.5557, 5.9384, 4.8252, 3.7451, 5.8703, 2.7885

To the best of your ability, determine what the mean and standard deviation are of this random variable. (This question is a bit open-ended, so there could be more than one correct way of justifying your answer.)

Exercise 13. Suppose I tell you that the following list of 20 numbers is a random sample from a Gaussian random variable, but I don't tell you the mean or standard deviation. Also, around one or two of the numbers was corrupted by noise, computational error, tabulation error, etc., so that it is totally unrelated to the actual Gaussian random variable.

$-1.2045, -1.4829, -0.3616, -0.3743, -2.7298, -1.0601, -1.3298, 0.2554, 6.1865, 1.2185$
$-2.7273, -0.8453, -3.4282, -3.2270, -1.0137, 2.0653, -5.5393, -0.2572, -1.4512, 1.2347$

To the best of your ability, determine what the mean and standard deviation are of this random variable. Supposing you had instead a billion numbers, and 5 or 10 percent of them were corrupted samples, can you come up with some automatic way of throwing out the corrupted samples? (Once again, there could be more than one right answer here; the question is intentionally open-ended.)

Exercise 14. Let $b_1, \ldots, b_n$ be distinct numbers, representing the quality of n people. Suppose n people arrive to interview for a job, one at a time, in a random order. That is, every possible arrival order of these people is equally likely. We can think of an arrival ordering of the people as an ordered list of the form $a_1, \ldots, a_n$, where the list $a_1, \ldots, a_n$ is a permutation of the numbers $b_1, \ldots, b_n$. Moreover, we interpret a_1 as the rank of the first person to arrive, a_2 as the rank of the second person to arrive, and so on. And all possible permutations of the numbers $b_1, \ldots, b_n$ are equally likely to occur.

For each $i \in \{1, \ldots, n\}$, upon interviewing the i^{th} person, if $a_i > a_j$ for all $1 \leq j < i$, then the i^{th} person is hired. That is, if the person currently being interviewed is better than the previous candidates, she will be hired. What is the expected number of hirings that will be made? (Hint: let $X_i = 1$ if the i^{th} person to arrive is hired, and let $X_i = 0$ otherwise. Consider $\sum_{i=1}^{n} X_i$.)