STOR 565 Homework 1

1. Let $X > 0$ be a positive, continuous random variable with density f_X. Use the CDF method to find the density of $Y = X^{-1}$ in terms of f_X.

2. Recall that the variance of a random variable X is defined by $\text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$. Carefully establish the following.

(a) If a, b are constants, then $\text{Var}(aX + b) = a^2 \, \text{Var}(X)$

(b) $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ (expand the square in the definition)

(c) $\mathbb{E}X^2 \geq (EX)^2$.

3. In this problem we find an upper bound on the variance of a random variable with values in a finite interval. Let X be a random variable taking values in the finite interval $[0, c]$. You may assume that X is discrete, though this is not necessary for this problem.

(a) Show that $\mathbb{E}X \leq c$ and $\mathbb{E}X^2 \leq c \, \mathbb{E}X$.

(b) Recall that $\text{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$. Use the inequalities above to show that

$$\text{Var}(X) \leq c^2[u(1-u)] \quad \text{where} \quad u = \frac{\mathbb{E}X}{c} \in [0, 1].$$

(c) Use this inequality and simple calculus to show that $\text{Var}(X) \leq c^2/4$ if $X \in [0, c]$.

(d) Use this result to show that if X is a random variable taking values in an interval $[a, b]$ with $-\infty < a < b < \infty$ then $\text{Var}(X) \leq (b-a)^2/4$

(e) It turns out that the general bound cannot be improved. To see this, show that the variance of the random variable $X \in [a, b]$ with $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 1/2$ is equal to the bound you found above.

4. Let $\langle x, y \rangle = x^t y = \sum_{i=1}^d x_i \, y_i$ be the usual inner product in $\mathbb{R}^d$. Recall that the norm of a vector $x \in \mathbb{R}^d$ is defined by $||x|| = \langle x, x \rangle^{1/2}$

(a) Show that $||x|| = 0$ if and only if $x = 0$.

(b) Use the definition of the norm to show that $||x + y||^2 = ||x||^2 + 2\langle x, y \rangle + ||y||^2$.

(c) Use this equation and the Cauchy Schwarz inequality to establish the triangle inequality for the vector norm, namely $||x + y|| \leq ||x|| + ||y||$.

(d) The standard Euclidean distance between two vectors $x, y \in \mathbb{R}^d$ is defined by $d(x, y) = ||x - y||$. Use part (c) to establish that $d(x, y) \leq d(x, z) + d(z, y)$ for any vectors $x, y, z \in \mathbb{R}^d$. Draw a picture illustrating this result.

5. Show that if $f(x)$ is bounded and $X \sim \text{Poiss}(\lambda)$ then $\mathbb{E}[\lambda f(X + 1)] = \mathbb{E}[Xf(X)]$. Here $\text{Poiss}(\lambda)$ denotes the usual Poisson distribution with pmf $p(k) = e^{-\lambda}\lambda^k/k!$ for $k \geq 0$.

6. Let $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ be two sequences of real numbers.

(a) Show that $\min\{a_i\} + \min\{b_i\} \ \leq \ \min\{a_i + b_i\} \ \leq \ \min\{a_i\} + \max\{b_i\}$.

Hints: For the first inequality, note that the leftmost term is less than or equal to $a_j + b_j$ for every j. For the second inequality, note that the middle term is less than or equal to $a_j + b_j$ where $a_j = \min\{a_i\}$.

(b) As clearly as you can, provide an English language explanation of the inequalities above.

(c) Following the arguments from the lecture, show that $\max\{-b_i\} = -\min\{b_i\}$.

(d) Use the results above to show that

$$\min\{a_i\} - \max\{b_i\} \ \leq \ \min\{a_i - b_i\} \ \leq \ \min\{a_i\} - \min\{b_i\}.$$

7. In each case below find $\min_{x \in \mathcal{X}} f(x)$, $\text{argmin}_{x \in \mathcal{X}} f(x)$, $\max_{x \in \mathcal{X}} f(x)$, and $\text{argmax}_{x \in \mathcal{X}} f(x)$. Indicate when the min or the max do not exist. It may help to sketch the functions.

(a) $f(x) = \sin x$ with $\mathcal{X} = [0, 2\pi]$ and $\mathcal{X} = [0, \pi]$

(b) $f(x) = \min(x^2, 1)$ with $\mathcal{X} = [0, 2]$ and $\mathcal{X} = (-2, 2]$

8. The probability that an individual has a certain rate disease is about 1 percent. If they have the disease, the chance that they test positive is 90 percent. If they do not have the disease, the chance that they nevertheless test positive is 9 percent. What is the probability that someone who tests positive actually has the disease? (Use Bayes Formula.) What does this say about the test?