Homework 5 (due Tuesday, June 11th at 11:59pm)

(Warm up: 1-10). These exercises are easy, and they serve the only purpose of making you think (and review) a bit more deeply about the basic properties of r.v.'s and expectation. I will provide no solutions for these exercises. Please come to office hour if you wish to discuss them.
(Hw: 11-16). These exercises are of standard difficulty.

1. (4 pts) Let the r.v. $X \in [0, b]$ have $E[X] = a$ with $0 < a < b$. Find upper bounds for

   - $P(X \geq c)$.
   - $P(X \leq c)$.

   for some $c \in (0, b)$.

2. (4 pts) Let the cdf of the r.v. X be $F_X(x) = 1 - e^{-\lambda x}$ for $\lambda \in (0, \infty)$. Find $f_X(x)$.

3. (4 pts) Let the cdf of the r.v. X be $F_X(x) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$ for $0 < k \leq n$. Find $p_X(x)$.

4. (4 pts) Let the pdf of the r.v. X be $f_X(x) = \min\{x, 2-x\} \mathbb{1}(0 \leq x \leq 2)$. Compute $P(1/4 \leq X \leq 5/4)$. *Remark:* given that I changed this exercise 3 times, you will get full score for any answer you provide (but not for a blank answer).

5. (4 pts) Let the pmf of the r.v. X be $p_X(x) = c\mathbb{1}_A(x)$ where $A = \{0, \frac{1}{100}, \frac{2}{100}, \ldots, \frac{99}{100}, 1\}$. Compute the normalizing constant c and $P(0.3 \leq X \leq 0.7)$.

6. (4 pts) Let the support of the r.v.'s X_1 and X_2 satisfy all of the following constraints:

   - $x_1^2 + x_2^2 \leq 2$
   - $\frac{1}{2}x_1 \leq x_2 \leq 2x_1$ and $x_1 \geq 0$ (a cone)
   - $x_2 \leq \frac{1}{3}$

   Draw the support on a 2D plane.

7. (4 pts) In class we said that

$$\frac{\partial \phi(\sqrt{y})}{\partial y} = \frac{1}{2\sqrt{y}} f_Y(\sqrt{y})$$

where f_Y is the pdf of the standard normal. Prove this claim.
Hint: you will need Leibniz rule for integration:

$$\frac{\partial}{\partial x}\int_{a(x)}^{b(x)} f(x,t)dt$$
$$= f(x,b(x))\frac{\partial}{\partial x}b(x) - f(x,a(x))\frac{\partial}{\partial x}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x}f(x,t)dt$$

8. (4 pts) Find the cdf of $Y = X^3$ where $X \sim \text{Uniform}(0,1)$.
Be careful to write the support carefully.

9. (4 pts) Let $a \in \mathbb{R}$ and X be some continuous r.v.. Prove the following properties using the pdf's:

   - $Cov(X,X) = V(X)$ (you can prove it without pdf's)
   - $Cov(aX,bY) = abCov(X,Y)$
   - $E[X_1X_2|X_2 = x_2] = x_2E[X_1|X_2]$ (trivial)
   - $E_{X_2}[E_{X_1}[X_1X_2X_3|X_2,X_3]|X_3] = x_3E_{X_2}[X_2E[X_1|X_2,X_3]|X_3]$

10. (4 pts) Let $X_1 \sim \text{Bernoulli}(p)$ and $X_2 \sim \text{Bernoulli}(f(|X_1 - 1/2|))$ for $f(y) = y^\alpha$ for $\alpha > 0$. Compute $E[X_2]$.

11. (10 pts) Consider independent trials, each of which results in outcome i, $i = 0,\dots,k$, with probability p_i, $\sum_{i=0}^k p_i = 1$. Let N be the number of trials needed to obtain an outcome that is not equal to 0 and let X be that outcome.

    (a) Find $P(N = n)$, $n \geq 1$.
    (b) Find $P(X = j)$, $j = 1,\dots,k$.
    (c) Are N and X independent?

    Solution:

    (a) Consider a successful trial as getting any nonzero element. Thus, the success chance is $p = 1 - p_0$. Notice that in this context, N is geometric(p), so $P(N = n) = (1 - p_0)p_0^{n-1}$.
    (b) Intuitively, X is more likely to be values corresponding to higher p_i. In fact, we are just rescaling the p_i's to exclude zero.

Let M be the distribution of any trial, so $P(M = m) = p_m$ for $m = 0, \dots, k$. Then:
$P(X = j) = P(M = j|M \neq 0) = \frac{P(M=j \cap M \neq 0)}{P(M \neq 0)} = \frac{p_j}{1-p_0}$
Note that this is equivalent to $\frac{p_j}{p_1+p_2+\ldots+p_k}$.

(c) Yes, they are independent. When the first nonzero occurs, and which specific nonzero value it is, are independent, because knowing that the value is nonzero on the first trial vs. on the 20^{th} trial doesn't tell you anything other than it being nonzero.

12. (10 pts) Let $X_1, X_2 \overset{iid}{\sim}$ Uniform(0,1). [1] Let $Y = X_1 + X_2$. Compute $F_Y(y) = P(X_1 + X_2 \leq y)$.
Hint: follow very closely what has been done in Quiz 3, exercise 3.

Solution:
Following the same steps as in the the proof in the quiz, you should obtain the following cdf.

$$F_Y(y) = \begin{cases} 0 \text{ if } y < 0 \\ \frac{y^2}{2} \text{ if } 0 \leq y \leq 1 \\ -\frac{y^2}{2} + 2y - 1 \text{ if } 1 < y \leq 2 \\ 1 \text{ o/w} \end{cases}$$

13. (10 pts) Let X, Y, Z be three continuous r.v.'s. Prove the following statements using the pdf's:

(a) $X \perp\!\!\!\perp (Y, Z)$ implies $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Z$

(b) $X \perp\!\!\!\perp (Y, Z)$ implies $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$

(c) $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z$ imply together that $X \perp\!\!\!\perp (Y, Z)$

Notice that the assumption of continuity is just to make the notation easier.

Solution:
Let f_Q be the pdf of the r.v. Q. Let's start from the first statement.

[1] X_1 and X_2 are two independent and identically distributed r.v.'s with Uniform between between 0 and 1, that is

$$f_X(x) = \mathbb{1}(x \in [0, 1]).$$

(a) For the first part of the implication,

$$f_{X,Y}(x,y) = \int f_{X,Y,Z}(x,y,z)dz$$
$$= \int f_X(x) f_{Y,Z}(y,z)dz = f_X(x)f_Y(y)\int f_{Z|Y=y}(z)dz$$
$$= f_X(x)f_Y(y)$$

The proof for the second part follows exactly the same strategy.

(b) For the first part of the implication,

$$f_{X,Y|Z}(x,y) = f_{X|(Y,Z)}(x)f_{Y|Z}(y)$$
$$= f_X(x)f_{Y|Z}(y) = f_{X|Z}(x)f_{Y|Z}(y)$$

where in the last equality we used the implication of the first statement. The second part of the implication follows the same proof.

(c)

$$f_{X,Y,Z}(x,y,z) = f_{X|Y,Z}(x)f_{Y,Z}(y,z)$$
$$= f_{X|Z}(x)f_{Y,Z}(y,z) = f_X(x)f_{Y,Z}(y,z)$$

14. (10 pts) Let $X \sim \text{Beta}(\alpha, \beta)$ and $Y \sim \text{Bernoulli}(X)$ with $y \in \{0,1\}$ being its realization. Compute the pdf of X conditional on Y, that is $p_{X|Y=y}$.

Solution:
In this exercise we are going *Bayesian*!
By Bayes theorem we know that

$$p_{X|Y=y}(x) = \frac{p_{Y|X=x}(y)p_X(x)}{p_Y(y)} \propto p_{Y|X=x}(y)p_X(x)$$

since the denominator is just a normalizing constant. Now,

$$p_{Y|X=x}(y)p_X(x) \propto x^y(1-x)^{1-y}x^{\alpha-1}(1-x)^{\beta-1}$$
$$= x^{\alpha+y-1}(1-x)^{\beta-y}.$$

Therefore we obtained the kernel of the distribution, and we recognize that this probability density function belongs to the family of Beta distributions. Consequently we can just conclude that

$$X|Y = y \sim \text{Beta}(\alpha + y, \beta + 1 - y).$$

15. (10 pts) For two r.v.'s X, Y, prove that

$$V(Y) = E[V(Y|X)] + V(E[Y|X]).$$

Solution:

$$\begin{aligned} V(Y) &= E[(Y - E[Y])^2] \\ &= E[(Y - E[Y|X] + E[Y|X] - E[Y])^2] \\ &= E[V(Y|X)] - 2E[(Y - E[Y|X])(E[Y] - E[Y|X])] + E[(E[Y|X] - E[Y])^2] \end{aligned}$$

Let's focus on the last term.

$$E[(E[Y|X] - E[Y])^2] = E[(E[Y|X] - E[E[Y|X]])^2] = V(E[Y|X])$$

by the tower property of expectation. Now, we also need to prove that the term in the middle is equal to 0. This is a bit more tricky, but for the same reason you can see that

$$\begin{aligned} &E[(Y - E[Y|X])(E[Y] - E[Y|X])] \\ &= E[Y]^2 + E[E[Y|X]^2] - E[YE[Y|X]] - E[E[Y]E[Y|X]] \\ &= E[Y]^2 + E[E[Y|X]^2] - E[E[Y|X]E[Y|X]] - E[Y]E[E[Y|X]] \\ &= E[Y^2] + E[E[Y|X]^2] - E[E[Y|X]^2] - E[Y]^2 = 0. \end{aligned}$$

16. (10 pts) We are going to prove a useful inequality, I'll guide you through the steps.
Let $X \in [x_l, x_m]$.

(a) Find the minimizer t of $h(t) = E[(X - t)^2]$.
Hint: you can either take the first two derivatives - the second to check that it's a minimum - or add/subtract $E[X]$.

(b) Conclude the fact above that

$$V(X) \leq h\left(\frac{x_l + x_m}{2}\right).$$

(c) Prove that

$$h\left(\frac{x_l + x_m}{2}\right) \leq \frac{1}{4}E[((X - x_l) - (X - x_m))^2]$$

Hint: remember that $X + Y \leq X - Y$ for $X \geq 0$ and $Y \leq 0$.

(d) Finally conclude that

$$V(X) \leq \frac{(x_m - x_l)^2}{4}.$$

Solution:
This is known as Popoviciu's inequality! Let's prove it.

- Taking the derivative:

$$\frac{\partial h(t)}{\partial t} = -2E[X] + 2t = 0 \implies t = E[X]$$
$$\frac{\partial^2 h(t)}{\partial t^2} = 2 > 0$$

  Alternatively,

$$E[(X - E[X] + E[X] - t)^2] = E[(X - E[X])^2] + (E[X] - t)^2$$

  where the first term does not depend on t, hence the minimum will be given by the minimization of the second term. This occurs when $t = E[X]$.
- $V(X) = h(E[X]) \leq h((x_m + x_l)/2)$.
- Using the hint, we have

$$E\left[(X - \frac{x_l + x_m)}{2})^2\right] = \frac{1}{4}E[((X - x_l) + (X - x_m))^2]$$
$$\leq \frac{1}{4}E[((X - x_l) - (X - x_m))^2] = \frac{1}{4}(x_m - x_l)^2$$

- By the results in the second and third bullet points, we now have

$$V(X) \leq \frac{1}{4}(x_m - x_l)^2.$$