## Sample Midterm Solutions

1. *Short questions*

   a. No, $\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^2 \leq \mathrm{E}(X-2)^2 \leq (7-2)^2 = 25$.

   b. No, a binomial r.v. has mean $np$ which is necessarily larger than its variance $np(1-p)$.

   c. $\mathrm{Var}\,(2X - 2Y) = 4\mathrm{Var}\,(X) + 4\mathrm{Var}\,(-Y) = 4\mathrm{Var}\,(X) + 4\mathrm{Var}\,(Y) = 4a$.

   d. 0.5. The sum of Gaussian random variables is Gaussian, so $4X + 3Y - Z - W$ is a Gaussian centered at zero.

   e. Split up by independence: $\mathrm{E}(X^2) = \int_0^2 x^2 \frac{1}{2} dx = 2$ and $\mathrm{E}(e^{2Y}) = \int_0^\infty e^{2y} 3 e^{-3y} dy = 3$.

   f. $\mathrm{P}(A) = \int_0^1 x^2 dx = 1/3$.

   g. $\mathrm{E}(X) = 0$, $\mathrm{E}(XY) = (-1)(1)\frac{1}{3} + (0)(0)\frac{1}{3} + (1)(1)\frac{1}{3} = 0$ yes uncorrelated.

   h. Not independent $\mathrm{P}(X = 1, Y = 1) = 1/3 \neq \mathrm{P}(X = 1)\mathrm{P}(Y = 1) = \frac{1}{3}\frac{2}{3} = 2/9$.

   i. $\mathrm{E}\left((X_n - 1)^2\right) = 1/n \to 0$, s $X_n$ converges to 1 *in mean square*.

   j. No. For example take a constant sequence of $X \in \{0, 2\}$ with equal probability.

   k. Convergence in distribution does not imply convergence in probability. Also the random variables do not have to be defined on the same probability space.

   l. $\mathrm{E}\left(\mathrm{E}\left(\frac{1}{f_X(X)} \middle| Y\right)\right) = \mathrm{E}\left(\frac{1}{f_X(X)}\right) = \int_1^3 \frac{1}{f_X(X)} f_X(X)\, dx = \int_1^3 dx = 2$.

2. *Monty Hall problem*

   a. The sample space for the cars and goats is: {(car, goat, goat), (goat, car, goat), (goat, goat, car)}. The $\sigma$-algebra is the power set of the sample space. The probability measure is uniform over the sample space.

   b. Yes! We are not modeling the initial decision probabilistically.

   $$\mathrm{P}(\text{Win if no switch}) = \mathrm{P}(\text{Car behind chosen door}) = \frac{1}{3}, \qquad (1)$$

   $$\mathrm{P}(\text{Win if switch}) = \mathrm{P}(\text{Car not behind chosen door}) = \frac{2}{3}. \qquad (2)$$

   c. By Bayes' theorem, if $S$ represents switching and $W$ represents winning,

   $$\mathrm{P}(S|W) = \frac{P(S)\,P(W|S)}{\mathrm{P}(S)P(W|S) + \mathrm{P}(S^c)P(W|S^c)} = \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3}} = \frac{2}{3}. \qquad (3)$$

   d. Let us assume that we stick to the switching strategy. The probability of winning the goat is $1/3$ and this is the stopping criterion, so we have a geometric random variable with

parameter 1/3 (and consequently mean 2).

$$E(\text{cars}) = \sum_{k=1}^{\infty} (k-1)\, p_K\,(k) \tag{4}$$

$$= \sum_{k=1}^{\infty} k p_K\,(k) - \sum_{k=1}^{\infty} p_K\,(k) \tag{5}$$

$$= E\,(K) - 1 = 2. \tag{6}$$

3. *Family*

   a. $X$ be the number of children in a given family, its range are the integers from 0 to $\infty$. The family will stop when they have two female children. Therefore $X$ can be expressed as a sum of two geometric random variables $X = T_1 + T_2$ where $T_i$ is geometric with parameter $p$. $E(T_1) = \sum k(1-p)^{k-1}p = 1/p$, so $E(X) = 2/p$.

   b. Let $R$ be the number of red haired kids. $E(R) = E(E(R|X)) = E(qX) = 2q/p$ since given $X = n$ each $n$ kid has the same chance of having red hair.

   c. To obtain the probability of having $x$ kids we add the probability of the disjoint events

   $$\{\text{girl, boy, boy, \ldots, boy, girl}\} \tag{7}$$
   $$\{\text{boy, girl, boy, \ldots, boy, girl}\} \tag{8}$$
   $$\ldots \tag{9}$$
   $$\{\text{boy, boy, boy, \ldots, girl, girl}\} \tag{10}$$
   $$\tag{11}$$

   There are $x - 1$ such events and each of them has probability $p^2\,(1-p)^{x-2}$. Since $R$ and $X$ are independent

   $$p_{X,R}\,(x,1) = (x-1)\,p^2\,(1-p)^{x-2} \cdot x\,(1-q)^{x-1}\,q, \tag{12}$$

   $$p_X\,(1) = \sum_{x=2}^{\infty} x\,(x-1)\,p^2\,(1-p)^{x-2}\,(1-q)\,q. \tag{13}$$

   We conclude

   $$p_{X|R}(x|1) = \frac{x\,(x-1)\,p^2\,(1-p)^{x-2}\,(1-q)^{x-1}\,q}{\sum_{u=2}^{\infty} u\,(u-1)\,p^2\,(1-p)^{u-2}\,(1-q)^{u-1}\,q}. \tag{14}$$

4. *Late*

   a. $X$ and $Y$ be the arrival times of Peter and Paula measured in hours, and $R = 1$ if rains and 0 else. $f_{X,Y} = \frac{1}{3}\frac{1}{2}\frac{1}{2}|_{\text{on } [11,13]^2} + \frac{2}{3}1 \cdot 1|_{\text{on } [11.5,12.5]^2}$, so it is 3/4 on the inner square and 1/12 on the outer squre minus the inner.

   b. $P(X > 12, Y < 12) + P(X < 12, Y > 12) = 1/2$.

   c. $P(R = 1|X > 12, Y > 12) = \frac{P(X>12,Y>12|R=1)P(R=1)}{P(X>12,Y>12)} = \frac{1/4 \cdot 1/3}{1/4} = 1/3$ which doesn't change anything because no matter what the distributions are centered around the same mean. If it rains they are precautious to leave early to keep the average arrival time the same.

d. $E(XY) = \int \int xy f_{X,Y} dx dy = (13^2 - 11^2)^2/48 + (12.5^2 - 11.5^2)^2/6 = 144$ so their covariance is zero. They are uncorrelated.

e. Not independent, $f_X = \int f_{X,Y} dy = \int_{11}^{13} \frac{1}{12} dy + \int_{11.5}^{12.5} \frac{2}{3} dy = \frac{1}{6}|_{\text{on}[11,13]} + \frac{2}{3}|_{\text{on}[11.5,12.5]}$ so the marginal of $X$ is 5/6 on [11.5,12.5] and 1/6 on [11,11.5] and [12.5,13]. By symmetry its the same for $Y$, their product is not equal to the joint density. It makes sense because if we know that one arrived at 12:55pm then it must have rained, which immediately constrains the distribution of the arrival time of the other person.

5. *Overbooking*

   a. Let $T$ be the number of tickets sold, and $P$ is the number of passengers who show up. Then given $T = n$, $P$ is binomial$(n, 0.9)$ so $\text{Var}(P|T = n) = n0.9 \cdot 0.1$.

   b. For $T = n$, we use the Central Limit Theorem as in Problem 1 of Homework 5, $P(P \leq 300) = P(\frac{P - 0.9n}{\sqrt{0.09n}} \leq \frac{300 - 0.9n}{0.3\sqrt{n}}) > 95\%$. Roughly $n$ should be $Q^{-1}(0.05) \approx 1.65 = \frac{300 - 0.9n}{0.3\sqrt{n}}$, solving for $n$ gives $n \leq 324$.

   c. The probability of a particular passenger being left out

   $$P(\text{passenger i left out}) = P(\text{someone is left out}) P(\text{passenger i left out} \mathbin{-} \text{someone is left out})$$
   $$= (0.9^{301}) \, 300. \tag{15}$$

   d. The gain is $301 \cdot 600$ if nobody is left out, and $301 \cdot 600 - 2600$ if someone is left out.

   $$E(\text{Gain}) = 301 \cdot 600 - P(\text{someone is left out}) \, 2600 = 301 \cdot 600 - 0.9^{301} 2600 \approx 180600.$$

   The probability of someone being left out is negligible.

6. *Christmas*

   a. We can create a Bernoulli variable (though you are free to assign any value you like): Let $B$ be 1 if turkey is cooked, and 0 if chicken is cooked. Then we know that $P(B = 1) = 0.3$ and $P(B = 0) = 0.7$, and $E(B) = 0.3$ & $\text{Var}(B) = 0.21$. Let $Y$ be the cooking time measured in hours. This is a random variable with an unknown distribution but we know its (conditional) properties: $E(Y|B = 1) = 4$, $E(Y|B = 0) = 2$. From this we can compute: $E(Y) = E(E(Y|B)) = 4 \cdot 0.3 + 2 \cdot 0.7 = 2.6$.

   b. By Markov's inequality: $P(Y > 5) \leq \frac{2.6}{5} = 52\%$.

   c. $\text{Var}(Y|X = 1) = 0.5$, $\text{Var}(Y|X = 0) = 0.25$ so $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)) = (0.5 \cdot 0.3 + 0.25 \cdot 0.7) + (4^2 \cdot 0.3 + (2)^2 \cdot 0.7 - (4 \cdot 0.3 + 2 \cdot 0.7)^2) = 1.165$. By Chebyshev's inequality: $P(Y > 5) \leq P(|Y - 2.6| > 2.4) \leq \frac{1}{5} \approx 20\%$.

   d. $E(BY) = E(E(BY|B)) = E(B E(Y|B)) = 1 \cdot 4 \cdot 0.3 + 0 \cdot 2.5 \cdot 0.7$, therefore $\text{Cov}(B, Y) = 1.2 - (0.3 \cdot 2.95) = 0.315$, and correlation is $\frac{0.315}{\sqrt{0.21 \cdot 0.797}} \approx 0.77$. Surprise! Cooking turkey and increased cooking time are positively correlated.