**Stat 512**
**Midterm Examination**
**November 10, 2015**

**Closed book, closed notes.**

**You must stop work precisely when you are told time is up. No paper will be accepted from a student still writing after the class has been told the exam is over.**

**In addition to the theorems covered in class, you may use the following facts without proof:**

- If $X \sim \mathcal{E}(\lambda)$ (an exponential random variable with density $f_X(x) = \lambda e^{-\lambda x}\mathbf{1}_{[0<x<\infty]}$ for some $\lambda > 0$), then $EX = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$.
- If $X \sim \mathcal{N}(\mu, \sigma^2)$ is a normal random variable, then $EX = \mu$, $Var(X) = \sigma^2$, and the third and fourth central moments are 0, and $3\sigma^4$, respectively.

1. (35 points) Let $X$ be a random variable having density function for some constant $a > 0$

$$f_X(x) = a(x-4)^2\mathbf{1}_{[1<x<4]}.$$

(a) Find $a$.
**Ans:** Because the density must integrate to 1:

$$\begin{aligned}\int_1^4 f_X(x)\,dx &= a\int_1^4 (x-4)^2\,dx \\ &= a\left[\frac{(x-4)^3}{3}\right]_1^4 = a\left[0 - \frac{-27}{3}\right] \\ &= 9a \\ &\Rightarrow a = \frac{1}{9}\end{aligned}$$

(b) Find the cumulative distribution function $F_X(x)$.
**Ans:** Integrating the density for $1 \le x \le 4$:

$$F_X(x) = \int_1^x f_X(u)\,du = \frac{1}{9}\int_1^x (u-4)^2\,dx = \frac{1}{9}\left[\frac{(x-4)^3}{3} + 9\right] = \frac{(x-4)^3}{27} + 1,$$

so we can write the cdf as

$$F_X(x) = \frac{(x-4)^3}{27}\mathbf{1}_{[1,4)}(x) + \mathbf{1}_{[1,\infty)}(x)$$

which expands to (though I find the above form easier)

$$F_X(x) = \left(\frac{x^3}{27} - \frac{4x^2}{9} + \frac{16x}{9} - \frac{37}{27}\right)\mathbf{1}_{[1,4)}(x) + \mathbf{1}_{[4,\infty)}(x).$$

(c) Find $E[X]$.

**Ans:** Integrating

$$
\begin{aligned}
E[X] &= \frac{1}{9}\int_1^4 x(x-4)^2\,dx = \frac{1}{9}\int_1^4 \left[x^3 - 8x^2 + 16x\right]\,dx \\
&= \frac{1}{9}\left[\frac{x^4}{4} - \frac{8x^3}{3} + 8x^2\right]_1^4 \\
&= \frac{1}{9}\left[64 - \frac{512}{3} + 128\right] - \frac{1}{9}\left[\frac{1}{4} - \frac{8}{3} + 8\right] \\
&= \frac{1}{9}\left[\frac{64}{3} - \frac{67}{12}\right] = \frac{21}{12} = \frac{7}{4}
\end{aligned}
$$

(d) Find $Var(X)$.

**Ans:** Using the computational formula: $Var(X) = E[X^2] - E^2[X]$, we first find $E[X^2]$

$$
\begin{aligned}
E[X^2] &= \frac{1}{9}\int_1^4 x^2(x-4)^2\,dx = \frac{1}{9}\int_1^4 \left[x^4 - 8x^3 + 16x^2\right]\,dx \\
&= \frac{1}{9}\left[\frac{x^5}{5} - 2x^4 + \frac{16x^3}{3}\right]_1^4 \\
&= \frac{1}{9}\left[\frac{1024}{5} - 512 + \frac{1024}{3}\right] - \frac{1}{9}\left[\frac{1}{5} - 2 + \frac{16}{3}\right] \\
&= \frac{1}{9}\left[\frac{1023}{5} - 510 + \frac{1008}{3}\right] = \frac{1}{9}\left[\frac{459}{15}\right] = \frac{51}{15} = \frac{17}{5} \\
Var(X) &= \frac{17}{5} - \left(\frac{7}{4}\right)^2 = \frac{272}{80} - \frac{245}{80} = \frac{27}{80}
\end{aligned}
$$

(e) Let $Y = 1/X$. Find the cumulative distribution function $F_Y(y)$.

**Ans:** Using the cdf for $X$,

$$
\begin{aligned}
Pr[Y \leq y] &= Pr\left[\frac{1}{X} \leq y\right] = Pr\left[\frac{1}{y} \leq X\right] = 1 - Pr\left[X \leq \frac{1}{y}\right] \\
&= 1 - \frac{\left(\frac{1}{y} - 4\right)^3}{27}\mathbf{1}_{[1,4)}\left(\frac{1}{y}\right) - \mathbf{1}_{[4,\infty)}\left(\frac{1}{y}\right) \\
&= 1 + \frac{(4y-1)^3}{27y^3}\mathbf{1}_{(0.25,1)}(y) - \mathbf{1}_{(-\infty,0.25)}(y) \\
&= \frac{(4y-1)^3}{27y^3}\mathbf{1}_{(0.25,1)}(y) + \mathbf{1}_{[1,\infty)}(y)
\end{aligned}
$$

which expands to

$$
F_Y(y) = \left(\frac{64}{27} - \frac{16}{9y} + \frac{4}{9y^2} - \frac{1}{27y^3}\right)\mathbf{1}_{(0.25,1)}(y) + \mathbf{1}_{[1,\infty)}(y).
$$

(f) Find $E[Y]$.

**Ans:** Integrating to find $E[Y] = E\left[\frac{1}{X}\right]$ (I think this easier than using $f_Y(y)$)

$$\begin{aligned}
E[Y] &= \frac{1}{9}\int_1^4 \frac{(x-4)^2}{x}\,dx = \frac{1}{9}\int_1^4 \left[x - 8 + \frac{16}{x}\right]\,dx \\
&= \frac{1}{9}\left[\frac{x^2}{2} - 8x + 16\log(x)\right]_1^4 \\
&= \frac{1}{9}\left[8 - 32 + 16\log(4)\right] - \frac{1}{9}\left[\frac{1}{2} - 8 + 0\right] \\
&= \left[\frac{32\log(4) - 33}{18}\right]
\end{aligned}$$

(g) Find $Var(Y)$.

**Ans:** Using the computational formula: $Var(Y) = E\left[\frac{1}{X^2}\right] - E^2\left[\frac{1}{X}\right]$, we first find $E\left[\frac{1}{X^2}\right]$

$$\begin{aligned}
E[Y^2] &= \frac{1}{9}\int_1^4 \frac{(x-4)^2}{x^2}\,dx = \frac{1}{9}\int_1^4 \left[1 - \frac{8}{x} + \frac{16}{x^2}\right]\,dx \\
&= \frac{1}{9}\left[x - 8\log(x) - \frac{16}{x}\right]_1^4 \\
&= \frac{1}{9}\left[4 - 8\log(4) - 4\right] - \frac{1}{9}\left[1 - 0 - 16\right] = \left[\frac{15 - 8\log(4)}{9}\right] \\
Var(Y) &= \left[\frac{15 - 8\log(4)}{9}\right] - \left[\frac{32\log(4) - 33}{18}\right]^2 \\
&= \left[\frac{270 - 288\log(4)}{324}\right] - \left[\frac{1024\log^2(4) - 2112\log(4) + 1089}{324}\right]^2 \\
&= \left[\frac{-549 + 1824\log(4) - 1024\log^2(4)}{324}\right]
\end{aligned}$$

2. (25 points) Let $X, Y$ be random variables having density function

$$f_{X,Y}(x,y) = ye^{-x}\mathbf{1}_{[0<y<x]}\mathbf{1}_{[0<x<\infty]}.$$

(a) Are $X$ and $Y$ independent? Very briefly explain your reasoning.

**Ans:** No, the density cannot be factored into terms involving only $x$ and only $y$.

(b) Find the marginal density of $X$, $f_X(x)$.

**Ans:** Integrate the joint density over $y$

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^x ye^{-x}\mathbf{1}_{[0<x<\infty]}\,dy = e^{-x}\mathbf{1}_{[0<x<\infty]}\int_0^x y\,dy \\
&= e^{-x}\mathbf{1}_{[0<x<\infty]}\left[\frac{y^2}{2}\right]_0^x = \frac{1}{2}x^2e^{-x}\mathbf{1}_{[0<x<\infty]}
\end{aligned}$$

(So $X \sim \Gamma(2,1,0)$, but you did not need to state that.)

(c) Find the marginal density of $Y$, $f_Y(y)$.
**Ans:** Integrate the joint density over $x$, noting that $\mathbf{1}_{[0<y<x]}\mathbf{1}_{[0<x<\infty]} = \mathbf{1}_{[0<y<\infty]}\mathbf{1}_{[y<x<\infty]}$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_y^{\infty} ye^{-x}\mathbf{1}_{[0<y<\infty]}\,dx = y\mathbf{1}_{[0<y<\infty]}\int_y^{\infty} e^{-x}\,dx \\ &= y\mathbf{1}_{[0<y<\infty]}\left[-e^{-x}\right]_y^{\infty} = ye^{-y}\mathbf{1}_{[0<y<\infty]} \end{aligned}$$

(So $Y \sim \Gamma(1,1,0)$, but you did not need to state that.)

(d) Find the conditional density of $Y$ given $X = x$, $f_{Y|X}(y|x)$.
**Ans:** Using the formula for the conditional density for $x \in (0,\infty)$

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{ye^{-x}\mathbf{1}_{[0<y<x]}\mathbf{1}_{[0<x<\infty]}}{\frac{1}{2}x^2e^{-x}\mathbf{1}_{[0<x<\infty]}} \\ &= \frac{2}{x^2}y\mathbf{1}_{[0<y<x]} \end{aligned}$$

(e) Find $E[Y \mid X = x]$.
**Ans:** For $x \in (0,\infty)$

$$E[Y \mid X = x] = \int_{-\infty}^{\infty} y\,f_{Y|X}(y|x)\,dy = \int_0^x \frac{2}{x^2}y^2\,dy = \left[\frac{2}{3x^2}y^3\right]_0^x = \frac{2}{3}x.$$

3. (20 points) The chi-squared distribution with $n$ degrees of freedom is derived as the distribution of the sum

$$\chi_n^2 = \sum_{i=1}^n Z_i^2,$$

where $Z_1, \ldots, Z_n$ are independent, identically distributed standard normal random variables (so $Z_i \sim \mathcal{N}(0,1)$). Suppose random variable $V \sim \chi_n^2$.

(a) Derive $E[V]$.
**Ans:** We use the definition of the chi square distribution by considering the sum $S$ of $n$ independent squared standard normal random variables $S = \sum_{i=1}^n Z_i^2$. From the computational formula, $Var(Z_i) = E[Z_i^2] - E^2[Z_i]$, so $E[Z_i^2] = Var(Z_i) + E^2[Z_i] = 1 + 0 = 1$. Then, by linearity of integration

$$E[S] = E\left[\sum_{i=1}^n Z_i^2\right] = \sum_{i=1}^n E[Z_i^2] = n.$$

And because $V$ has the same distribution as $S$, $E[V] = n$.

(b) Derive $Var(V)$.

**Ans:** Again considering the distribution of $S$ defined in part a, by independence and identical distribution we know

$$Var(S) = Var\left(\sum_{i=1}^{n} Z_i^2\right) = \sum_{i=1}^{n} Var(Z_i^2) = nVar(Z_i^2).$$

From the computational formula, $Var(Z_i^2) = E[Z_i^4] - E^2[Z_i^2]$. Now, the kurtosis for this mean 0 random variable is

$$E[(Z_i - E[Z_i])^4] = E[(Z_i - 0)^4] = E[Z_i^4].$$

Further, because the kurtosis of a $\mathcal{N}(\mu, \sigma^2)$ random variable is $3\sigma^4$ and $Z_i \sim \mathcal{N}(0,1)$, we know $kurt(Z_i) = 3$ and thus $E[Z_i^4] = 3$. So we have by independence $Var(Z_i^2) = E[Z_i^4] - E^2[Z_i^2] = 3 - 1^2 = 2$ and

$$Var(S) = nVar(Z_i^2) = 2n.$$

Because $V$ has the same distribution as $S$, $Var(V) = 2n$.

4. (55 points) Let $X_1, X_2, X_n$ be independent, identically distributed random variables with density

$$f_{X_i}(x) = \frac{2x}{\theta^2}\mathbf{1}_{(0,\theta)}(x).$$

(a) Find $E[\,X_i\,]$ and $Var[\,X_i\,)$.

**Ans:**

$$\begin{aligned}
E[X_i] &= \int_{-\infty}^{\infty} x\, f_{X_i}(x)\, dx = \int_0^{\theta} \frac{2}{\theta^2} x^2\, dx = \left[\frac{2}{3\theta^2}x^3\right]_0^{\theta} = \frac{2}{3}\theta \\
E[X_i^2] &= \int_{-\infty}^{\infty} x^2\, f_{X_i}(x)\, dx = \int_0^{\theta} \frac{2}{\theta^2} x^3\, dx = \left[\frac{2}{4\theta^2}x^4\right]_0^{\theta} = \frac{1}{2}\theta^2 \\
Var(X_i) &= E[X_i^2] - E^2[X_i] = \frac{1}{2}\theta^2 - \left(\frac{2}{3}\theta\right) = \frac{\theta^2}{18}
\end{aligned}$$

So we have

$$X_i \sim \left(\frac{2\theta}{3}, \frac{\theta^2}{18}\right).$$

(b) For $W = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, find $E[W]$ and $Var(W)$.

**Ans:** By linearity of integration, we know that for $n$ independent identically distributed random variables having $X_i \sim (\mu, \sigma^2)$, the sample mean $\overline{X}$ has distribution $\overline{X} \sim (\mu, \frac{\sigma^2}{n})$. So

$$W \sim \left(\frac{2\theta}{3}, \frac{\theta^2}{18n}\right).$$

(c) For what value of $a$ would "estimator" $\hat{\theta}_W = aW$ be unbiased (i.e., have $E[\hat{\theta}_W] = \theta$)?
**Ans:** Now $E[aW] = aE[W] = \frac{2a}{3}\theta$. So if we want $E[aW] = \theta$, we find

$$\begin{aligned} E[\hat{\theta}_W] = E[aW] &= \frac{2a}{3}\theta = \theta \\ \Rightarrow \qquad a &= \frac{3}{2}. \end{aligned}$$

(d) What is $Var(\hat{\theta}_W)$ for that value of $a$?
**Ans:**

$$\begin{aligned} Var(\hat{\theta}_W) &= Var(aW) = a^2 Var(W) \\ &= \frac{9}{4} Var(W) = \frac{9}{4}\frac{\theta^2}{18n} \\ &= \frac{\theta^2}{8n}. \end{aligned}$$

(e) For $Y = \max(X_1, X_2, \ldots, X_n)$, find the cumulative distribution function $F_Y(y \,|\, \theta)$.
**Ans:**

$$\begin{aligned} F_Y(y) &= Pr(Y \le y) = Pr\left(\max\{X_1, \ldots, X_n\} \le y\right) = Pr\left(\cap_{i=1}^n \{X_i \le y\}\right) \\ &= \prod_{i=1}^n Pr\left(X_i \le y\right) \qquad \text{(by independence)} \\ &= \left[Pr\left(X_i \le y\right)\right]^n \qquad \text{(by identical distribution)} \\ &= \left[\int_{-\infty}^y f_{X_i}(x)\, dx\right]^n \\ &= \left[\mathbf{1}_{(0,\theta)}(y) \int_{-\infty}^y \frac{2}{\theta^2} x\, dx + \mathbf{1}_{[\theta,\infty)}(y) \int_{-\infty}^\theta \frac{2}{\theta^2} x\, dx\right]^n \\ &= \mathbf{1}_{(0,\theta)}(y) \left[\frac{y^2}{\theta^2}\right]^n + \mathbf{1}_{[\theta,\infty)}(y) \\ &= \frac{y^{2n}}{\theta^{2n}} \mathbf{1}_{(0,\theta)}(y) + \mathbf{1}_{[\theta,\infty)}(y) \end{aligned}$$

(f) For $Y = \max(X_1, X_2, \ldots, X_n)$, find the probability density function $f_Y(y \,|\, \theta)$.
**Ans:**

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left[\frac{y^{2n}}{\theta^{2n}} \mathbf{1}_{(0,\theta)}(y) + \mathbf{1}_{[\theta,\infty)}(y)\right] \\ &= \frac{2n\, y^{2n-1}}{\theta^{2n}} \mathbf{1}_{(0,\theta)}(y) \end{aligned}$$

(g) Find $E[Y]$.
**Ans:**

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_0^{\theta} \frac{2n\, y^{2n}}{\theta^{2n}}\, dy = \left[\frac{2n}{2n+1}\frac{y^{2n+1}}{\theta^{2n}}\right]_0^{\theta}$$
$$= \frac{2n}{2n+1}\theta$$

(h) Find $Var(Y)$.
**Ans:**

$$E[Y^2] = \int_{-\infty}^{\infty} y^2 f_Y(y)\, dy = \int_0^{\theta} \frac{2n\, y^{2n+1}}{\theta^{2n}}\, dy = \left[\frac{2n}{2n+2}\frac{y^{2n+2}}{\theta^{2n}}\right]_0^{\theta}$$
$$= \frac{2n}{2n+2}\theta^2$$
$$Var(Y) = E[Y^2] - E^2[Y] = \frac{2n}{2n+2}\theta^2 - \left[\frac{2n}{2n+1}\theta\right]^2$$
$$= 2n\theta^2 \left[\frac{(2n+1)^2 - 2n(2n+2)}{(2n+2)(2n+1)^2}\right]$$
$$= \frac{2n}{(2n+2)(2n+1)^2}\theta^2 = \frac{n}{(n+1)(2n+1)^2}\theta^2$$

(i) For what value of $b$ would “estimator” $\hat{\theta}_Y = bY$ be unbiased (i.e., have $E[\hat{\theta}_Y] = \theta$?
**Ans:** Now $E[bY] = bE[Y] = \frac{2n}{2n+1}b\theta$. So if we want $E[bY] = \theta$, we find

$$E[\hat{\theta}_Y] = E[bY] = \frac{2n}{2n+1}b\theta = \theta$$
$$\Rightarrow \qquad b = \frac{2n+1}{2n}.$$

(j) What is $Var(\hat{\theta}_Y)$ for that value of $b$?
**Ans:**

$$Var(\hat{\theta}_Y) = Var(bY) = b^2 Var(Y) = \frac{(2n+1)^2}{4n^2} Var(Y)$$
$$= \frac{(2n+1)^2}{4n^2}\,\frac{2n}{(2n+2)(2n+1)^2}\theta^2$$
$$= \frac{\theta^2}{2n(2n+2)} = \frac{\theta^2}{4n(n+1)}.$$

(k) Which of the unbiased estimators $\hat{\theta}_W$ and $\hat{\theta}_Y$ is more precise (i.e., has lower variance)?
**Ans:** We can consider the ratio

$$\frac{Var(\hat{\theta}_W)}{Var(\hat{\theta}_Y)} = \frac{\frac{\theta^2}{8n}}{\frac{\theta^2}{4n(n+1)}}$$

$$= \frac{n+1}{2}$$

Hence, the variance of $\hat{\theta}_W$ is greater than the variance of $\hat{\theta}_Y$ for all $n > 1$, and the "relative efficiency" of $\hat{\theta}_Y$ to $\hat{\theta}_W$ (which this ratio measures) grows in proportion to $n$ with increasing sample size.
*Note that when $n = 1$, $W = Y$ and all the above formulas are in perfect agreement with each other.*

5. (10 points) Suppose normally distributed random variable $Y \sim \mathcal{N}(\theta, \tau^2)$, and suppose that random variable $X$ is conditionally normally distributed according to

$$(X|Y = y) \sim \mathcal{N}(y, y^2).$$

(a) Find $E[X]$, the unconditional expectation of $X$.
**Ans:** Using the double expectation formula

$$E[X] = E_Y\,[\,E[X\,|\,Y]\,] = E_Y[\,Y\,] = \theta.$$

(b) Find $Var[X]$, the unconditional variance of $X$.
**Ans:** We have

$$\begin{aligned}Var(X) &= E_Y\,[\,Var(X|Y)\,] = Var_Y(\,E[X|Y]\,)\\ &= E_Y[\,Y^2\,] + Var_Y(Y)\\ &= (Var(Y) + E^2(Y)) + Var(Y)\\ &= \tau^2 + \theta^2 + \tau^2\\ &= 2\tau^2 + \theta^2\end{aligned}$$

6. (35 points) Let $Y \sim \mathcal{B}(1, p_Y)$ and $X \sim \mathcal{B}(1, p_X)$. Further suppose

$$Pr[Y = 1, X = 1] = \theta p_Y p_X.$$

(a) Find $\rho = corr(Y, X)$.
**Ans:** We consider $p_X \in (0, 1)$ and $p_Y \in (0, 1)$, as the possibility that either probability is 0 or 1 is not particularly interesting.
Noting that the expectation of a binary variable is just the probability that that variable is 1, we use the computational formula for covariance and the definition of correlation to

find

$$\begin{aligned}
Cov(Y,X) &= E[XY] - E[Y]\,E[X] \\
&= \theta p_Y p_X - p_Y p_X \\
&= p_Y p_X(\theta - 1) \\
\rho = corr(Y,X) &= \frac{Cov(Y,X)}{\sqrt{Var(Y)\,Var(X)}} \\
&= \frac{p_Y p_X(\theta-1)}{\sqrt{p_Y(1-p_Y)p_X(1-p_X)}} = (\theta-1)\sqrt{\frac{p_Y}{(1-p_Y)}\frac{p_X}{(1-p_X)}}
\end{aligned}$$

(b) What values of $\theta$ are valid?
**Ans:** As a probability, we must have

$$0 \le \theta p_Y p_X \le 1.$$

We also know that for any two events $A$ and $B$, $Pr(AB) \le \min(Pr(A), Pr(B)) \le 1$. So

$$0 \le \theta p_Y p_X \le \min(p_Y, p_X).$$

Hence,

$$0 \le \theta \le \frac{\min(p_Y, p_X)}{p_Y p_X} = \frac{1}{\max(p_Y, p_X)} = \min\left(\frac{1}{p_Y}, \frac{1}{p_X}\right).$$

We also consider the other "cells" of the 2 x 2 table of $Pr(Y = y, X = X)$:

| | $x = 0$ | $x = 1$ | $Pr(Y = y)$ |
|---|---|---|---|
| $y = 0$ | $1 - p_Y - p_X + \theta p_Y p_X$ | $p_X(1 - \theta p_Y)$ | $1 - p_Y$ |
| $y = 1$ | $p_Y(1 - \theta p_X)$ | $\theta p_Y p_X$ | $p_Y$ |
| $Pr(X = x)$ | $1 - p_X$ | $p_X$ | |

They must all be probabilities. It is sufficient to consider

$$Pr(Y = 0, X = 0) = 1 - p_Y - p_X + \theta p_Y p_X.$$

So we also have

$$0 \le 1 - p_Y - p_X + \theta p_Y p_X \le \min(1 - p_y, 1 - p_X)$$

which yields

$$\frac{p_X + p_Y - 1}{p_Y p_X} \le \theta \le \min\left(\frac{1}{p_Y}, \frac{1}{p_X}\right),$$

and combining these results we get

$$\max\left(0, \frac{p_X + p_Y - 1}{p_Y p_X}\right) \le \theta \le \min\left(\frac{1}{p_Y}, \frac{1}{p_X}\right),$$

(c) When is $\rho = 1$?
**Ans:** For any specified $p_Y$ and $p_X$, correlation $\rho$ is maximized by

$$\theta = \frac{1}{\max(p_Y, p_X)} = \min\left(\frac{1}{p_Y}, \frac{1}{p_X}\right),$$

in which case

$$\begin{aligned}\rho &= (\theta - 1)\sqrt{\frac{p_Y}{(1-p_Y)}\frac{p_X}{(1-p_X)}} \\ &= \left(\frac{1}{\max(p_Y, p_X)} - 1\right)\sqrt{\frac{p_Y}{(1-p_Y)}\frac{p_X}{(1-p_X)}} \\ &= \left(\frac{1-\max(p_Y, p_X)}{\max(p_Y, p_X)}\right)\sqrt{\frac{p_Y}{(1-p_Y)}\frac{p_X}{(1-p_X)}} \\ &= \sqrt{\left(\frac{1-\max(p_Y, p_X)}{\max(p_Y, p_X)}\right)\left(\frac{\min(p_Y, p_X)}{1-\min(p_Y, p_X)}\right)}\end{aligned}$$

and $\rho = 1$ only if $\max(p_Y, p_X) = \min(p_Y, p_X)$, (or more succinctly, $p_Y = p_X$) <u>and</u> $\theta = \frac{1}{p_Y}$. Hence, for these binary random variables, $\rho = 1$ can only be true if

$$Pr(Y = 1, X = 1) = Pr(Y = 1) = Pr(X = 1)$$

which in turn is only true if if $Y = X$.

(d) When is $\rho = 0$? Does $\rho = 0$ imply independence in this setting? Very briefly explain your reasoning.
**Ans:** By inspection, $\rho = 0$ for $p_Y, p_X \in (0, 1)$ if and only if $\theta = 1$. Note that for $\theta = 1$

$$Pr(Y = 1, X = 1) = Pr(Y = 1)Pr(X = 1).$$

and we have that events $\{Y = 1\}$ and $\{X = 1\}$ are independent. By properties of independent events, we thus know that complements of independent events are also independent, so

$$\begin{aligned}Pr(Y = 1, X = 0) &= Pr(Y = 1)Pr(X = 0) \\ Pr(Y = 0, X = 1) &= Pr(Y = 0)Pr(X = 1) \\ Pr(Y = 0, X = 0) &= Pr(Y = 0)Pr(X = 0)\end{aligned}$$

Hence, for $\forall x, y \in \{0, 1\}$, $Pr(Y = y, X = x) = Pr(Y = y)Pr(X = x)$, and uncorrelated binary variables $X$ and $Y$ are independent.

(e) What is the minimum value of $\rho$ as a function of $p_Y$ and $p_X$.
**Ans:** Because $\rho$ is linear in $\theta$, we know the minimum value of $\rho$ will be attained at the minimum value of $\theta$, which is

$$\theta = \max\left(0, \frac{p_X + p_Y - 1}{p_Y p_X}\right).$$

Thus for $p_X + p_Y \leq 1$

$$\rho \geq -\sqrt{\frac{p_Y}{(1-p_Y)}\frac{p_X}{(1-p_X)}}.$$

For $p_X + p_Y > 1$

$$\begin{aligned}\rho &\geq \left(\frac{p_X + p_Y - 1}{p_X p_Y} - 1\right)\sqrt{\frac{p_Y}{(1-p_Y)}\frac{p_X}{(1-p_X)}} \\ &\geq -\sqrt{\frac{(1-p_Y)}{p_Y}\frac{(1-p_X)}{p_X}}\end{aligned}$$

Note that in either case this only attains the general lower bound of -1 for $\rho$ when $Y$ and $X$ are such that $p_Y = 1 - p_X$, which includes $p_X = p_Y = \frac{1}{2}$. In this setting, $\theta = 0$, thus, the most negative correlation of $\rho = -1$ demands $X = 1 - Y$.
*(Note we will later use the Cauchy-Schwarz inequality to prove that corr$(X, Y) = 1$ only if $Y = aX + b$ for some $a > 0$, and corr$(X, Y) = -1$ only if $Y = aX + b$ for some $a < 0$.)*

(f) Find $Var(Y + X)$.
**Ans:**

$$\begin{aligned}Var(Y + X) &= Var(Y) + Var(X) + 2Cov(Y, X) \\ &= p_Y(1-p_Y) + p_X(1-p_X) + 2(\theta - 1)p_Y p_X \\ &= (p_Y + p_X)(1 - (p_Y + p_X)) + 2\theta p_Y p_X\end{aligned}$$

(g) Find $Var(Y - X)$.
**Ans:**

$$\begin{aligned}Var(Y - X) &= Var(Y) + Var(X) - 2Cov(Y, X) \\ &= p_Y(1-p_Y) + p_X(1-p_X) - 2(\theta - 1)p_Y p_X\end{aligned}$$