Question 1: (25 Points) For a certain species of trees, the diameter (D) and usable height (H) of each tree was measured (in feet). The following probability distribution is estimated.

V

8	10
12.5	15.625
18	22.5
24.5	30.625

	H	
D	20	25
1.0	0.16	0.09
1.25	0.15	0.30
1.50	0.03	0.17
1.75	0	0.10

0.25
0.45
0.20
0.10

0.34 0.66

Since a tree trunk is roughly a cylinder, the volume V of usable wood that it contains is given approximately by:

$$V = 0.4D^2H$$

a) Are diameter and height independent? Do not make use of Covariance to answer this problem.

Is $P(D=1 \text{ and } H=20) \overset{?}{=} P(D=1)P(H=20)$

$= 0.25(0.34) = 0.085 \neq 0.16$

So, H and D are not independent.

b) Calculate the average volume of a tree E(V).

$V = 0.4D^2H$ $\quad E(V) = \sum_{i=1}^{8} V_i P_V(v_i)$

$= 8(0.16) + 10(0.09) + 12.5(0.15)$

$+ 15.625(0.30) + 18(0.03) + 22.5(0.17)$

$+ 24.5(0) + 30.625(0.10)$

$= 1.28 + 0.9 + 1.875 + 4.6875$

$+ 0.54 + 3.825 + 3.0625 = 16.17$

c) Write an expression for ~~E~~ V$(H - 5D)$ in terms of ~~expectations~~ variances of H, D and the covariance of H and D. Do not calculate.

$Var(H-5D) = Var(H) + Var(5D) - 2\,Cov(H, 5D)$

$= Var(H) + 25\,Var(D) - 2(5)\,Cov(H, D)$

$= Var(H) + 25\,Var(D) - 10\,Cov(H, D)$

d) Calculate P(D=1.25|H=20).

$$\frac{P(D=1.25, H=20)}{P(H=20)} = \frac{0.15}{0.34} = 0.441$$

Question 2: (30 Points)

a) Find the pdf of $W = \sqrt{Y}$ if Y has the pdf $f_Y(y) = 2y$ for $0 \le y \le 1$.

$$F_W(y) = P(W \le y) = P(\sqrt{Y} \le y) = P(Y \le y^2) = F_Y(y^2) \quad 0 \le y \le 1.$$

$$f_W(y) = F_W'(y) = \frac{d}{dy} F_Y(y^2) = 2y\, f_Y(y^2)$$

$$\Rightarrow \quad f_W(y) = 2y\,(2y^2) = 4y^3 \qquad 0 \le y \le 1$$

b) Let $Y_1, \dots, Y_{48}$ be a random sample from the exponential density $f_Y(y) = \lambda e^{-\lambda y}, y > 0$. Approximate $P(Y_1 + \dots + Y_{48} < 25/\lambda)$. (Hint: $E(Y) = 1/\lambda$ and $V(Y) = 1/\lambda^2$)

$$P(Y_1 + \dots + Y_{48} < 25/\lambda) = P\left(\bar{Y} < \frac{25}{48\lambda}\right)$$

$$= P\left(\bar{Y} < \frac{0.52}{\lambda}\right)$$

$$\cong P\left(Z < \frac{0.52/\lambda - 1/\lambda}{\frac{1/\lambda}{\sqrt{48}}}\right)$$

$$= P(Z < -3.32) \cong 0$$

c) Let $Y_1, \dots, Y_n$ be a random sample from the pdf $f_Y(y) = \frac{2y}{\theta^2}, 0 \le y \le \theta$. Find an unbiased estimator for θ. (Hint: Consider $\hat{\theta} = \bar{Y}$)

$$E[\bar{Y}] = E[Y] = \int_0^\theta y \frac{2y}{\theta^2}\, dy = \frac{2}{\theta^2} \frac{y^3}{3} \Big|_0^\theta$$

$$= \frac{2}{3}\theta$$

So $\hat{\theta} = \frac{3}{2}\bar{Y}$ is an unbiased estimator for θ.

(because $E[3/2\, \bar{Y}] = \theta$)

Question 3: (25 Points) A "planet transit" is a rare event in which a planet appears to cross in front of its star as seen from Earth. Assume that the number of planet transits discovered for every 3,000 stars follows a Poisson distribution with $\lambda = 5$ as announced by NASA.

a) What is the probability that more than 10 planet transits will be seen, in the next 3,000 stars monitored by NASA? Use Normal approximation to Poisson to answer this question.

$$P(X>10) = P\left(\frac{X-\lambda}{\sqrt{\lambda}} > \frac{10+0.5-\lambda}{\sqrt{\lambda}}\right) = P\left(Z > \frac{10.5-5}{\sqrt{5}}\right)$$

$$= P\left(Z > \frac{5.5}{2.236} = 2.459\right)$$

$$= P(Z > 2.459) = 1 - 0.9931$$
$$= 0.0069$$

b) It is also reported by NASA that 20% of the stars that the astrophysicists observe are from our galaxy Milky Way. Among the 3,000 stars monitored, what is the approximate probability that at least 550 stars are from Milky Way?

$p = 0.20 \quad n = 3000 \quad \frac{X - np}{\sqrt{npq}} \Rightarrow \quad np = 3000.(0.20) = 600$

$npq = 480 \Rightarrow \sqrt{npq} \cong 21.9$

$P)\ P(X \geq 550)$

$$= P\left(\frac{X-np}{\sqrt{npq}} \geq \frac{550-0.5-600}{21.90}\right) = P\left(Z > \frac{-50.5}{21.90} = -2.30\right)$$

$$= P(Z > -2.30) = 0.9893$$

c) The time it takes to observe 3,000 stars is normally distributed with mean 1 year (=365days). If the probability that the next mission of NASA to observe 3,000 stars takes more than 400 days is 0.0764, find the standard deviation of the distribution.

$\mu = 365 \qquad P(X > 400) = 0.0764$

$P(Z > z) = 0.0764 = 1 - P(Z \leq z)$

$\Rightarrow P(Z \leq z) = 0.9236$

$P(X > 400) = P\left(\frac{X-\mu}{\sigma} > \frac{400-365}{\sigma} = 1.43\right) \qquad z = 1.43$

$400 - 365 = 14.3(\sigma) \Rightarrow \quad 14.3(\sigma) = 35 \qquad \sigma = \frac{35}{14.3} = 2.44$

Question 4: (30 Points) Consider the geometric distribution with parameter *p*.

a) Derive the maximum likelihood estimator for *p*.

$$L(p,X) = \ln \prod_{i=1}^{n} p(x_i,p) = \ln \prod_{i=1}^{n} (1-p)^{x_i-1} p = \ln (1-p)^{\sum x_i - n} p^n$$

$$= (\sum x_i - n) \ln(1-p) + n \ln p$$

$$\Rightarrow \quad \frac{d\, L(p,X)}{dp} = -\frac{\sum x_i - n}{1-p} + \frac{n}{p} = 0$$

$$\frac{n}{p} = \frac{\sum x_i - n}{1-p}$$

$$n - np = p \sum x_i - np$$

$$n = p \sum x_i \Rightarrow \hat{p} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

b) Suppose the following observations come from a geometric distribution: 3, 2, 2, 1, 4. Estimate the parameter *p* using the MLE you found in part a).

$$\bar{x} = \frac{3+2+2+1+4}{5} = \frac{12}{5} = 2.4 \qquad \hat{p} = \frac{1}{2.4} = 0.41$$

c) Find the sample variance s^2 using the data in part b).

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \frac{(9+4+4+1+16) - 5(2.4)^2}{5-1}$$

$$= \frac{34 - 28.8}{4} = 1.3$$

d) In a problem of estimating the mean μ from a large sample, the investigators believe that the standard deviation is around 5.6 and they wish to estimate μ with an error margin of at most 0.7 and 98% confidence. What should the smallest sample size *n* be?

$$P\left(\bar{X} - z\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\frac{\sigma}{\sqrt{n}}\right) = 98\% \qquad z_{0.99} = 2.32$$

two sided

$$z\frac{\sigma}{\sqrt{n}} \leq 0.7 \Rightarrow 2.33\frac{5.6}{\sqrt{n}} \leq 0.7$$

$$\sqrt{n} \geq \frac{2.33(5.6)}{0.7} = 18.64$$

$$n \geq 347.4496 \Rightarrow n \geq 348$$