Problem IX.1: Let $X$ be uniformly distributed in $[0, 1]$. Given $X$, the random variable $Y$ is $B(N, X)$. 
(That is, $Y$’s distribution is binomial with parameters $N$ and $X$.)

a. Find the MLE $\hat{X}$ of $X$ given $Y$. 
b. Find the value of $N$ required so that $\sqrt{E((X - \hat{X})^2)} \leq 0.01E(X)$.

a. The MLE $\hat{X}$ is the value of $X$ that makes the observed $Y$ most likely. Thus we seek the value of $x$ such that $P(Y \mid X = x)$ is maximized. Note that the conditioning event $\{X = x\}$ occurs with probability zero. Recall that when we write the conditional probability of an event condition on the event $\{X = x\}$ we are using short hand for writing the limit of the conditional probability of an event conditioned on the event $\{x \leq X \leq x + \epsilon\}$ as $\epsilon$ goes to zero.

$$P(Y \mid X = x) = \binom{N}{Y} x^Y (1 - x)^{N-Y} \quad (1)$$

Differentiating this function with respect to $x$ and setting the result equal to zero we have the following for when $Y$ is neither 0 nor $N$.

$$\binom{N}{Y} \frac{\partial}{\partial x} (x^Y (1 - x)^{N-Y}) = 0$$
$$Y x^{Y-1} (1 - x)^{N-Y} - x^Y (N - Y)(1 - x)^{N-Y-1} = 0$$
$$Y(1 - x) = x(N - Y)$$
$$x = \frac{Y}{N} \quad (2)$$

Note that $x$ cannot be 1 or 0 since $Y$ is neither 0 nor $N$ respectively. Thus there is no division by zero in the third line.

When $Y = 0$, $P(Y \mid X = x) = (1 - x)^N$ and is maximized when $x = 0$. Similarly when $Y = N$, $P(Y \mid X = x) = x^N$ and is maximized when $x = 1$. Thus for any value of observed $Y$

$$\hat{X} = \frac{Y}{N} \quad (3)$$

This is answer is very reasonable since it is just the sample mean of a sequence of i.i.d. Bernoulli random variables.

b.
\[ E[(X - \hat{X})^2] = E[X^2 - \frac{2}{N}XY + \frac{Y^2}{N^2}] \]
\[ = E[X^2] - \frac{2}{N}E[XY] + \frac{1}{N^2}E[Y^2] \]
\[ = \frac{1}{3} - \frac{2}{N}E[E[XY \mid X]] + \frac{1}{N^2}E[E[Y^2 \mid X]] \]
\[ = \frac{1}{3} - \frac{2}{N}E[XE[Y \mid X]] + \frac{1}{N^2}E[NX(1 - X) + N^2X^2] \]
\[ = \frac{1}{3} - \frac{2}{N}E[XNX] + \frac{1}{N^2}E[NX + N(N - 1)X^2] \]
\[ = \frac{1}{3} - 2E[X^2] + \frac{1}{N}(E[X] + (N - 1)E[X^2]) \]
\[ = \frac{1}{3} - \frac{2}{3} + \frac{1}{2N} + \frac{1}{3} - \frac{1}{3N} \]
\[ = \frac{1}{6N} \]

\[ \sqrt{E[(X - \hat{X})^2]} \leq \frac{1}{100}E[X] \]
\[ \frac{1}{6N} \leq \frac{1}{40000} \]
\[ N \geq \frac{40000}{6} \]
\[ N \geq 6667 \]  

**Problem IX.2:** Let \( X, Y \) be independent and uniformly distributed in \([0, 1]\). Find the LLSE of \((X + Y)^n\) given \(X\).

\[ L[(X + Y)^n] = E[(X + Y)^n] + \frac{cov(X, (X + Y)^n)}{var(X)}(X - E[X]) \]  
\[ E[X] = 0.5 \text{ and } var(X) = \frac{1}{12}. \]

\[ E[(X + Y)^n] = E[\sum_{k=0}^{n} \binom{n}{k}X^kY^{n-k}] \]
\[ = \sum_{k=0}^{n} \binom{n}{k} \frac{1}{k + 1} \frac{1}{n - k + 1} \]
\[\text{cov}(X, (X + Y)^n) = E[X(X + Y)^n] - E[X]E[(X + Y)^n]\]
\[= E\left[\sum_{k=0}^{n} \binom{n}{k} X^{k+1}Y^{n-k}\right] - E[X]E[(X + Y)^n]\]
\[= \sum_{k=0}^{n} \binom{n}{k} \frac{1}{k+2} \frac{1}{n-k+1} - \frac{1}{2} \sum_{k=0}^{n} \binom{n}{k} \frac{1}{k+1} \frac{1}{n-k+1}\]  
\[= \sum_{k=0}^{n} \binom{n}{k} \frac{1}{n-k+1} \left(\frac{1}{k+2} - \frac{1}{2k+2}\right) \tag{8}\]

**Problem IX.3:** Let \(X, Y, Z\) be independent and standard Gaussian. Find the LLSE of \((X + 2Y)^4\) given \(X + Y + Z\).

Note that \(X + 2Y\) is distributed \(N(0, 5)\). Thus \(X + Y = \sqrt{5}U\) in distribution where \(U\) is distributed \(N(0, 1)\).

\[E[(X + 2Y)^4] = 25E[U^4]\]
\[= (25)(3)\]
\[= 75 \tag{9}\]

\[\text{cov}((X + 2Y)^4, X + Y + Z) = E[(X + 2Y)^4(X + Y + Z)] - E[(X + 2Y)^2]E[X + Y + Z]\]
\[= E[(X + 2Y)^4(X + 2Y - Y)]\]
\[= E[(X + 2Y)^5] - E[Y(X + 2Y)^4]\]
\[= - \sum_{i=0}^{4} \binom{4}{i} E[X^i](2^4-i)E[Y^{5-i}]\]
\[= 0 \tag{10}\]

Hence
\[L[(X + 2Y)^4] = 75 \tag{11}\]

**Problem IX.4:** Let \((X, Y)\) be the coordinates of a point chosen uniformly in the unit circle centered at \((2, 2)\). Find the LLSE of \(X\) given \(Y\).

Let \(X' = X - 2\) and \(Y' = Y - 2\). Then \((X', Y')\) will be uniform in the unit circle centered at \((0, 0)\). Note that \(E[X'] = E[Y'] = 0\) and \(E[X'Y'] = 0\) since \(X'\) and \(Y'\) are uncorrelated (see midterm II). Note also that \(\text{cov}(X, Y) = \text{cov}(X', Y') = 0\). Thus

\[L[X \mid Y] = E[X] + \frac{\text{cov}(X, Y)}{\text{var}(Y)}(Y - E[Y])\]
\[= 2 + 0(Y - E[Y])\]
\[= 2 \tag{12}\]
Problem IX.5: A digital transmission line introduces random bit errors. The line is used all the time and its transmission rate is $10^8$ bits per second. We assume that the errors are i.i.d. and that the bit error rate is $10^{-12}$. Estimate the time it takes to observe 200 errors.

Let $N$ denote the number of bits needed to observe 200 errors. Let $h_i = E[N \mid i \text{ errors have been observed}]$. Then for $i = 0, 1, \ldots, 199$

$$h_i = 1 + ph_{i+1} + (1 - p)h_i$$

and $h_{200} = 0$ where $p = 10^{-12}$.

With a little simplification, for $i = 0, 1, \ldots, 199$

$$h_i = \frac{1}{p} + h_{i+1}$$

This isn’t surprising since the amount of time to wait for the next error after observing an error is geometric and independent of the times when other errors occurred or will occur. This all follows from the errors being i.i.d.

Thus $h_0 = \frac{200}{10^{-12}}$. We need to scale this by time. Thus the average time it takes to observe 200 errors is $\frac{(200)(10^{12})}{10^8}$ seconds. Note that this is about 23 days!

Problem IX.6: Let $\{X_n, n \geq 1\}$ be i.i.d. random variables. Propose an unbiased estimator of the variance.

Unbiased means that the expected value of the estimator equals the expected value of the random variable being estimated. The random variable to be estimated is the variance - i.e. the random variable is a constant and hence always equal to its mean value for all $\omega \in \Omega$. We make a guess

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}{n - 1}$$

where $\overline{X}_n$ denotes the sample mean of $\{X_i, 1 \leq i \leq n\}$

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$$

We now show that this estimator is unbiased.

$$E[\hat{\sigma}_n^2] = \frac{1}{n - 1} \sum_{i=1}^{n} E[X_i^2] - 2X_i\overline{X}_n + (\overline{X}_n)^2$$

$$= \frac{1}{n - 1} (nE[X_i^2] - \frac{2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_iX_j] + \frac{n}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_iX_j])$$

$$= \frac{1}{n - 1} (nE[X_i^2] - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_iX_j])$$

$$= \frac{1}{n - 1} (nE[X_i^2] - \frac{1}{n} (nE[X_i^2] + n(n - 1)E[X_i]^2))$$

$$= \frac{1}{n - 1} ((n - 1)E[X_i^2] - (n - 1)E[X_i]^2)$$

$$= E[X_i^2] - E[X_i]^2$$

$$= var(X_i)$$
Problem IX.7: Let $X$ be uniformly distributed in $[0, 1]$. Given $X$, let $\{Y_n, n \geq 1\}$ be i.i.d. with $P[Y_n = 1|X] = X = 1 - P[Y_n = 0|X]$.  

a. Are the random variables $\{Y_n, n \geq 1\}$ independent? Explain.

b. Are these random variables positively correlated? Explain.

a. The random variables are not independent. It is sufficient to show that two distinct events involving $Y_i$’s are dependent. Consider the two events $\{Y_i = 1, 1 \leq i \leq n\}$ and $\{Y_i = 1, 1 \leq i \leq n+1\}$. These two events are the event that the first $n$ tosses are heads; the first $n+1$ tosses are heads. We showed in homework 6 problem 6 the following:

$$P(Y_i = 1, 1 \leq i \leq n+1 | Y_i = 1, 1 \leq i \leq n) = \frac{n+1}{n+2}$$

But

$$P(Y_i = 1, 1 \leq i \leq n+1) = E[1(Y_i = 1, 1 \leq i \leq n+1)]$$

$$= E[E[1(Y_i = 1, 1 \leq i \leq n+1) | X]]$$

$$= E[\prod_{i=1}^{n+1} E[1(Y_i = 1) | X]]$$

$$= E[E[1(Y_i = 1) | X]^{n+1}]$$

$$= E[X^{n+1}]$$

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

(19)

The two events are not independent thus the sequence of random variables is not independent. This shouldn’t be surprising. With each observation of $Y_i$ you get a better idea of what $X$ is which gives information about what the values of other $Y_i$ will be.

b. Positive correlation means that if for any arbitrary collection of $k$ $Y_i$ we observe that they tend to be 1 then another equally sized arbitrary collection of $Y_i$ will also tend to be 1. An analogous statement can be made when zeros are observed. We expect this sequence to be positively correlated by the reasoning given at the end of part a.

Positive correlation is established by showing that the covariance of an arbitrary sample of $k Y_i$ with another arbitrary sample of $k Y_i$ is positive. This is just a multivariate generalization of the scalar case. We consider the case where the sample sets are disjoint. The general case is not too much harder to show. Let $Y^{(1)}$ and $Y^{(2)}$ denote our arbitrary disjoint samples of $k Y_i$. Note these $Y_i$ need not occur in any order.

$$E[(Y^{(1)} - E[Y^{(1)}])(Y^{(2)} - E[Y^{(2)}])] = \sum_{i=1}^{k} E[Y_i^{(1)} Y_i^{(2)}] - E[Y_i^{(1)}] E[Y_i^{(2)}]$$

$$= k E[Y_i^{(1)} Y_i^{(2)}] - E[Y_i^{(1)}] E[Y_i^{(2)}]$$

(20)

The last step is possible since the two sets are disjoint.
\[ E[Y_i^{(1)} Y_i^{(2)}] - E[Y_i^{(1)}]E[Y_i^{(2)}] = E[E[Y_i^{(1)} Y_i^{(2)} | X]] - E[E[Y_i^{(1)} | X]]E[E[Y_i^{(2)} | X]] \\
= E[E[Y_i^{(1)} | X]^2] - E[E[Y_i^{(1)} | X]]^2 \\
= E[X^2] - E[X]^2 \\
= \frac{1}{12} \tag{21} \]

**Problem IX.8:** Let \( \{X_n, n \geq 1\} \) be i.i.d. and exponentially distributed with mean 1. Let \( \{Y_n, n \geq 1\} \) be i.i.d. and exponentially distributed with mean 2. You observe a sequence of random variables \( \{Z_n, n \geq 1\} \) that is equally likely to be \( \{X_n, n \geq 1\} \) or \( \{Y_n, n \geq 1\} \). How many random variables \( \{Z_1, \ldots, Z_n\} \) do you need to observe before being 95% confident of which sequence you are observing?

Let \( H_X \) denote the event that \( \{Z_n, n \geq 1\} = \{X_n, n \geq 1\} \) and \( H_Y \) denote the event that \( \{Z_n, n \geq 1\} = \{Y_n, n \geq 1\} \). We make the decision using MLE.

\[
\prod_{i=1}^{n} f_{Z_i}(Z_i | H_X) \frac{H_X}{H_Y} \prod_{i=1}^{n} f_{Z_i}(Z_i | H_Y) \\
\exp(-\sum_{i=1}^{n} Z_i) \frac{H_X}{H_Y} \frac{1}{2^n} \exp(-\sum_{i=1}^{n} \frac{Z_i}{2}) \\
\exp(-\sum_{i=1}^{n} \frac{Z_i}{2}) \frac{H_X}{H_Y} \frac{1}{2^n} \\
- \sum_{i=1}^{n} \frac{Z_i}{2} \frac{H_X}{H_Y} n \log \frac{1}{2} \\
\sum_{i=1}^{n} Z_i \frac{H_Y}{H_X} n \log 4 \tag{22} \]

To have 95% confidence is to have 95% probability of making the correct decision.

\[
P(\text{correct}) = P(\text{Decide}X | H_X)P(H_X) + P(\text{Decide}Y | H_Y)P(H_Y) \\
= P(\sum_{i=1}^{n} Z_i \leq n \log 4 | H_X) \frac{1}{2} + P(\sum_{i=1}^{n} Z_i > n \log 4 | H_Y) \frac{1}{2} \\
= P(\sum_{i=1}^{n} X_i \leq n \log 4) \frac{1}{2} + P(\sum_{i=1}^{n} Y_i > n \log 4) \frac{1}{2} \tag{23} \\
= \frac{1}{2} \int_{0}^{n \log 4} \frac{x^{n-1}e^{-x}}{(n-1)!} dx + \frac{1}{2} \int_{n \log 4}^{\infty} \frac{1}{2} \left( \frac{y}{2} \right)^{n-1} e^{-\frac{y}{2}} \frac{1}{(n-1)!} dy \]

Using your favorite numerical solver (I used Mathematica) we obtain confidence values for a few values of \( n \).
Problem IX.9: Construct a probability space and events \( A, B, C \) such that \( P(A \cap B) > P(A)P(B) \) and \( P(A \cap C) < P(A)P(C) \).

Recall that a probability space is a triple: \((\Omega, \mathcal{F}, P)\). The sigma algebra \( \mathcal{F} \) is the set of subsets of \( \Omega \) that have probability assignments.

Let \( \Omega = \{a, b, c\} \). Let \( \mathcal{F} = 2^\Omega \). Let \( P(\{a\}) = P(\{b\}) = P(\{c\}) = \frac{1}{3} \). Let \( A = \{a\}; B = \{a, b\}; \) and \( C = \{c\} \). Then \( P(A \cap B) = P(A) = \frac{1}{3} \) > \( \frac{2}{9} = P(A)P(B) \) and \( P(A \cap C) = 0 < \frac{1}{9} = P(A)P(C) \).

Problem IX.10: Give an example of random variables \( X \) and \( Y \) that are uncorrelated but not independent.

Consider \((X, Y)\) drawn uniformly from a unit disk centered at \((0, 0)\). \( X \) and \( Y \) are uncorrelated (See midterm II). \( X \) and \( Y \) are not independent however. To see this it is sufficient to show that for an event involving just \( X \) and an event involving just \( Y \) are not independent. Consider the events \( \{X \geq \frac{\sqrt{2}}{2}\} \) and \( \{Y \geq \frac{\sqrt{2}}{2}\} \).

\[
P(X \geq \frac{\sqrt{2}}{2} \mid Y \geq \frac{\sqrt{2}}{2}) = 0 \tag{24}
\]

But

\[
P(X \geq \frac{\sqrt{2}}{2}) = \frac{\pi}{4} - \frac{1}{2} \tag{25}
\]

Problem IX.11: Give an example of random variables \( X, Y, Z \) such that \( X \) and \( Y \) are independent given \( Z \) but \( P(X > Y) = 1 \).

Let \( X = 1, Y = 0, Z = \pi \). Clearly \( P(X > Y) = 1 \). Furthermore \( X, Y, \) and \( Z \) are independent since

\[
F_{X,Y,Z}(x, y, z) = 1(X \leq x, Y \leq y, Z \leq z) = 1(1 \leq x)1(0 \leq y)1(\pi \leq z) = F_X(x)F_Y(y)F_Z(z) \tag{26}
\]

Thus \( X \) and \( Y \) are independent conditioned on \( Z \) as well.

Problem IX.12: Let \( X, Y, Z \) be i.i.d. standard Gaussian random variables and let \( V = X + 2Y + 3Z \) and \( W = 3X + 2Y + Z \). Calculate \( E[X \mid V, W] \).

\( E[X \mid V, W] \) is an affine combination of \( V \) and \( W \).

\[
E[X \mid V, W] = a + b(V - E[V]) + c(W - E[W]) \tag{27}
\]
Note that

\[ E[E[X | V, W]] = a \]  \hspace{1cm} (28)

But the left hand side is just \( E[X] \) which is zero. Thus,

\[ E[X | V, W] = bV + cW \]  \hspace{1cm} (29)

where \( b \) and \( c \) are such that

\[ E[(X - bV - cW)V] = 0 \]
\[ E[(X - bV - cW)W] = 0 \]  \hspace{1cm} (30)

Simplifying the first orthogonality condition

\[
E[(X - bV - cW)V] = 0 \\
E[XV] - bE[V^2] - cE[VW] = 0 \\
1 - b(1 + 4 + 9) - c(3 + 4 + 3) = 0 \\
14b + 10c = 1 \]  \hspace{1cm} (31)

Simplifying the second orthogonality condition

\[
E[(X - bV - cW)W] = 0 \\
10b + 14c = 3 \]  \hspace{1cm} (32)

Solving this system of equations gives \( b = -\frac{1}{6} \) and \( c = \frac{1}{3} \). Thus

\[ E[X | V, W] = -\frac{1}{6}V + \frac{1}{3}W \]  \hspace{1cm} (33)