Problem VII.1: Let $X_1, X_2, \ldots$ be independent standard Gaussian random variables. Calculate

a. $E[X_1 + 2X_2 \mid 2X_1 + X_2]$.

b. $E[X_1 + 2X_2 + 3X_3 \mid X_1 + X_2 + X_3]$.

a. 

$$2(2X_1 + X_2) = E[4X_1 + 2X_2 \mid 2X_1 + X_2]$$

$$= E[3X_1 \mid 2X_1 + X_2] + E[X_1 + 2X_2 \mid 2X_1 + X_2]$$

Thus

$$E[X_1 + 2X_2 \mid 2X_1 + X_2] = 2(2X_1 + X_2) - 3E[X_1 \mid 2X_1 + X_2]$$

Note that $X_1$ and $2X_1 + X_2$ may be jointly defined. Thus $E[X_1 \mid 2X_1 + X_2]$ is an affine function of $2X_1 + X_2$. It must be zero mean since $E[E[X_1 \mid 2X_1 + X_2]] = E[X_1] = 0$. Thus,

$$E[X_1 \mid 2X_1 + X_2] = a(2X_1 + X_2)$$

To determine $a$ we apply the orthogonality property of conditional expectation.

$$E[(X_1 - a(2X_1 + X_2))(2X_1 + X_2)] = 0$$

$$2E[X_1^2] + E[X_1]E[X_2] = a(4E[X_1^2] + 4E[X_1]E[X_2] + E[X_2^2])$$

$$2 + 0 = a(4 + 0 + 1)$$

$$a = \frac{2}{5}$$

Thus, $E[X_1 \mid 2X_1 + X_2] = \frac{2}{5}(2X_1 + X_2)$

$$E[X_1 + 2X_2 \mid 2X_1 + X_2] = 2(2X_1 + X_2) - \frac{6}{5}(2X_1 + X_2)$$

$$= \frac{4}{5}(2X_1 + X_2)$$

b. Note that $X_1 + 2X_2 + 3X_3$ and $X_1 + X_2 + X_3$ may be jointly defined. Thus for some real $a$

$$E[X_1 + 2X_2 + 3X_3 \mid X_1 + X_2 + X_3] = a(X_1 + X_2 + X_3)$$

where $a$ is such that

$$E[(X_1 + 2X_2 + 3X_3 - a(X_1 + X_2 + X_3))(X_1 + X_2 + X_3)] = 0$$

$$E[X_1^2] + 2E[X_2^2] + 3E[X_3^2] = a(E[X_1^2] + E[X_2^2] + E[X_3^2])$$

$$\frac{6}{3}a$$

$$a = 2$$
Hence
\[ E[X_1 + 2X_2 + 3X_3 | X_1 + X_2 + X_3] = 2(X_1 + X_2 + X_3) \]  
\hspace{1cm} (8)

\textbf{Problem VII.2:} Let \( X \) be a \( N(\mu, \sigma^2) \) random variable. Calculate
\[ E((X + 1)^4). \]

Let \( Z = \frac{X - \mu}{\sigma} \). Then \( Z \) is a standard Gaussian random variable. You should check this for yourself. Let \( \zeta = \frac{\mu + 1}{\sigma} \). Then
\[
E[(X + 1)^4] = E[(\sigma Z + \sigma \zeta)^4] \\
= \sigma^4 E[(Z + \zeta)^4] \\
= \sigma^4 \left( \frac{4}{4} E[Z^4] + \frac{4}{3} E[Z^3] \zeta + \frac{4}{2} E[Z^2] \zeta^2 + \frac{4}{1} E[Z] \zeta^3 + \frac{4}{0} \zeta^4 \right) \\
= \sigma^4 (3 + 0 + 6\zeta^2 + 0 + \zeta^4) \\
= 3\sigma^4 + 6(\mu + 1)^2\sigma^2 + (\mu + 1)^4
\]

\[ E[X^4] = 3 \text{ when } X \text{ is standard Gaussian. For standard Gaussian random variables the odd moments} \]
\[ \text{are zero and the even moments are given by} \]
\[ E[X^{2m}] = \frac{(2m)!}{(2^m)(m!)} \]  
\hspace{1cm} (10)

This equation can be obtained through the characteristic function of a Gaussian random variable. You can also obtain it using the characteristic function of a random variable \( Y = X^2 \) where \( X \) is a standard Gaussian random variable.

\textbf{Problem VII.3:} Let \( X \) be a standard Gaussian random variable. Calculate
\[ E[X | (X - 1)^2]. \]

First suppose that \( X \) was distributed such that \( X > 1 \) with probability one - or almost surely. Then \( X - 1 > 0 \) almost surely. Then \( X = \sqrt{(X - 1)^2 + 1} = E[X | (X - 1)^2] \). If \( X \) was distributed such that \( X \leq 1 \) almost surely then \( X = -\sqrt{(X - 1)^2 + 1} = E[X | (X - 1)^2] \). We expect for a general distribution of \( X \), that the conditional expectation will be a convex combination of these two answers. Thus in general \( X \) can be expressed as the following function of \((X - 1)^2\).

\[ X = 1 + \sqrt{(X - 1)^2} \mathbb{1}(X \geq 1) - \sqrt{(X - 1)^2} \mathbb{1}(X < 1) \]  
\hspace{1cm} (11)

Let \( Z = X - 1 \). Then \( Z \) is distributed \( N(-1, 1) \). Thus,
\[
E[Z | Z^2] = \sqrt{Z^2} E[1(Z \geq 0) | Z^2] - \sqrt{Z^2} E[1(Z < 0) | Z^2] \\
= \sqrt{Z^2} P(Z \geq 0 | Z^2) - \sqrt{Z^2} P(Z < 0 | Z^2)
\]
\hspace{1cm} (12)

The following bit is presented informally. I will formulate this more rigorously in the near future.
\[ P(Z < 0 \mid Z^2 = z) = \frac{P(Z^2 = z, Z < 0)}{P(Z^2 = z)} = \frac{P(Z^2 = z, Z < 0)}{P(Z^2 = z, Z < 0) + P(Z^2 = z, Z \geq 0)} = \frac{P(Z = -\sqrt{z})}{P(Z = -\sqrt{z}) + P(Z = \sqrt{z})} = \frac{f_Z(-\sqrt{z})}{f_Z(-\sqrt{z}) + f_Z(\sqrt{z})} \]

Similarly,

\[ P(Z \geq 0 \mid Z^2 = z) = \frac{e^{-\frac{(\sqrt{z}+1)^2}{2}}}{e^{-\frac{(\sqrt{z}^2+1)^2}{2}} + e^{-\frac{(\sqrt{z}+1)^2}{2}}} \]

Substituting the above equalities

\[ E[Z \mid Z^2] = \sqrt{Z^2} \frac{e^{-\frac{(\sqrt{Z^2+1})^2}{2}}}{e^{-\frac{(\sqrt{Z^2+1})^2}{2}} + e^{-\frac{(\sqrt{Z^2})^2}{2}}} - \sqrt{Z^2} \frac{e^{-\frac{(\sqrt{Z^2+1})^2}{2}}}{e^{-\frac{(\sqrt{Z^2+1})^2}{2}} + e^{-\frac{(\sqrt{Z^2})^2}{2}}} \]

Thus,

\[ E[X \mid (X - 1)^2] = 1 + \sqrt{(X - 1)^2} \frac{e^{-\frac{(\sqrt{(X - 1)^2+1})^2}{2}}}{e^{-\frac{(\sqrt{(X - 1)^2+1})^2}{2}} + e^{-\frac{(\sqrt{(X - 1)^2+1})^2}{2}}} - \sqrt{(X - 1)^2} \frac{e^{-\frac{(\sqrt{(X - 1)^2+1})^2}{2}}}{e^{-\frac{(\sqrt{(X - 1)^2+1})^2}{2}} + e^{-\frac{(\sqrt{(X - 1)^2+1})^2}{2}}} \]

The following discussion about the Q-function can probably wait for another problem. :-) where

\[ Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \] This integral has no analytical closed form but is common enough that numerical approximations are very common. You can make your own Q and inverse Q function using MATLAB’s error function commands:

\[ Q(x) = \frac{1}{2} \text{erfc}(\frac{x}{\sqrt{2}}) \]

\[ Q^{-1}(y) = \sqrt{2} \text{erfinv}(1 - 2y) \]

**Problem VII.4:** Let \( X, Y \) be independent standard Gaussian random variables. Calculate \( P(3X + 4Y > 25) \).
Let $Z = 3X + 4Y$. Then $Z$ is distributed $N(0, 25)$ since the sum of Gaussian random variables is Gaussian with a mean that is the sum of the constituent means and with a variance that is the sum of the constituent variances.

$$P(Z > 25) = Q\left(\frac{25}{5}\right) = 2.8665e - 07$$ (18)

**Problem VII.5**: Let $X$ and $Y$ be independent standard Gaussian random variables. Calculate

$$E(|2X + 3Y - Z|^2)$$

where $Z = E[2X + 3Y | X + Y]$.

Note that this quantity is the mean square error of the conditional expectation estimate of $2X + 3Y$.

Let us first determine $Z$. By the orthogonality condition

$$E[(2X + 3Y - a(X + Y))(X + Y)] = 0$$

$$a = \frac{5}{2}$$

$$Z = \frac{5}{2}(X + Y)$$ (19)

$$E[|2X + 3Y - \frac{5}{2}(X + Y)|^2] = E[|\frac{Y - X}{2}|^2]$$

$$= \frac{1}{4}(1 + 1)$$

$$= \frac{1}{2}$$ (20)

**Problem VII.6**: Let $X, Y_1, Y_2, \ldots$ be independent standard Gaussian random variables and $a$ a positive real number. Calculate

$$E[X | aX + Y_1, aX + Y_2, \ldots, aX + Y_n]$$

Note that this conditional expectation will be an affine combination or the $aX + Y_i$ random variables. Since each of these and $X$ are zero mean, the affine combination will actually be a linear combination. Thus,

$$E[X | aX + Y_1, aX + Y_2, \ldots, aX + Y_n] = \sum_{j=1}^{n} b_j(aX + Y_j)$$ (21)

The orthogonality condition holds for every $aX + Y_i$ term. So for every $i$ between 1 and $n$
\[E[(X - \sum_{j=1}^{n} b_j(aX + Y_j))(aX + Y_i)] = 0\]

\[aE[X^2] + E[XY_i] = E[\sum_{j=1}^{n} b_j(aX + Y_j)(aX + Y_i)]\]

\[a = \sum_{j=1}^{n} b_jE[(aX + Y_j)(aX + Y_i)]\]

\[a = \sum_{j=1}^{n} b_j(a^2 E[X^2] + aE[XY_i] + aE[XY_j] + E[Y_i Y_j])\]

\[a = \sum_{j=1}^{n} b_j(a^2 + \delta_{ij})\]

\[a = a^2 \sum_{j=1}^{n} b_j + b_i\]

Note that the \(b_i\)'s are all the same by symmetry. Thus let \(b = b_i\).

\[a = a^2(nb) + b\]

\[b = \frac{a}{na^2 + 1}\] (23)

Thus,

\[E[X|aX + Y_1, aX + Y_2, \ldots, aX + Y_n] = \frac{a}{na^2 + 1} \sum_{j=1}^{n} (aX + Y_j)\] (24)

**Problem VII.7:** Let \(X\) be a standard Gaussian random variable and \(Y\) an independent random variable uniformly distributed in \(\{1, 2, \ldots, N\}\). Define \(Z = XY\). Calculate \(var(Z)\).

\[\text{var}(Z) = E[(XY - E[XY])^2]\]

\[= E[(XY - E[X]E[Y])^2]\]

\[= E[X^2Y^2] - 2E[X]E[Y]^2 + E[Y]^2\]

\[= E[X^2]E[Y^2]\]

\[= E[Y^2]\] (25)

\[= \sum_{k=1}^{N} \frac{k^2}{N}\]

\[= \frac{1 + 3N + 2N^2}{6}\]
Problem VII.8: Given $\mu$, the random variables $Y_1, Y_2, \ldots$ are independent and Gaussian with mean $\mu$ and variance $\sigma^2$. The mean $\mu$ is a random variable equal to $-1$ with probability 0.5 and is equal to $+1$ otherwise. We want to determine the number $N$ of observations $Y_1, \ldots, Y_N$ that allow us to determine whether $\mu = -1$ or $\mu = +1$ with a probability of error less than 1%.

We first need a decision rule for deciding which value $\mu$ takes on given that we observe $Y_1, \ldots, Y_N$. Since we are given the distribution of priors - i.e. the distribution of $\mu$ - we take the Bayes' approach.

Using Bayes' rule we can rewrite the above decision rule as

$$ f(Y_1, \ldots, Y_N | \mu = 1) P(\mu = 1) \geq \frac{f(Y_1, \ldots, Y_N | \mu = -1) P(\mu = -1)}{f(Y_1, \ldots, Y_N)} $$

(27)

We can make a few simplifications since there are terms common to both left and right sides.

$$ f(Y_1, \ldots, Y_N | \mu = 1) \geq \frac{f(Y_1, \ldots, Y_N | \mu = -1)}{f(Y_1, \ldots, Y_N)} $$

(28)

Note that conditioned on $\mu$, $Y_i$ are independent.

$$ \prod_{i=1}^{N} f(Y_i | \mu = 1) \geq \prod_{i=1}^{N} f(Y_i | \mu = -1) $$

(29)

Note that conditioned on $\mu = 1$, $Y_i$ is distributed $N(1, \sigma^2)$ and conditioned on $\mu = -1$, $Y$ is distributed $N(-1, \sigma^2)$.

$$ \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(Y_i - 1)^2}{2\sigma^2}} \geq \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(Y_i + 1)^2}{2\sigma^2}} $$

(30)

Making some simplifications:

$$ -\sum_{i=1}^{N} \frac{(Y_i - 1)^2}{2\sigma^2} \geq -\sum_{i=1}^{N} \frac{(Y_i + 1)^2}{2\sigma^2} $$

(31)

Making some more simplifications:

$$ -\sum_{i=1}^{N} \frac{(Y_i^2 - 2Y_i + 1)}{2\sigma^2} \geq -\sum_{i=1}^{N} (Y_i^2 + 2Y_i + 1) $$

(32)

$$ \sum_{i=1}^{N} Y_i \geq 0 $$

(33)

Let $X = \sum_{i=1}^{N} Y_i$. Then conditioned on $\mu$, $X$ is a Gaussian random variable with mean $N\mu$ and variance $N\sigma^2$. Let $\hat{\mu}_N$ denote our decision based on the above decision rule. Then the error event corresponds to $\hat{\mu}_N = 1$ conditioned on the event $\{\mu = -1\}$ or $\hat{\mu}_N = -1$ conditioned on the event $\{\mu = 1\}$. 
\( P(\text{error}) = P(\text{error} \mid \mu = -1)P(\mu = -1) + P(\text{error} \mid \mu = 1)P(\mu = 1) \)

\( = P(\hat{\mu} = 1 \mid \mu = -1)P(\mu = -1) + P(\hat{\mu} = -1 \mid \mu = 1)P(\mu = 1) \)

\( = \frac{1}{2}P(X \geq 0 \mid \mu = -1) + \frac{1}{2}P(X < 0 \mid \mu = 1) \)

\( = \frac{1}{2} \int_0^\infty \frac{1}{\sqrt{2\pi}N\sigma^2} e^{-\frac{(u+N)^2}{2N\sigma^2}} du + \frac{1}{2} \int_0^6 \frac{1}{\sqrt{2\pi}N\sigma^2} e^{-\frac{(u-N)^2}{2N\sigma^2}} du \)

\( \quad \quad \quad \quad \quad = \frac{1}{2} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du + \frac{1}{2} \int_0^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \)

\( = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \)

\( = Q\left(\sqrt{\frac{N}{\sigma}}\right) \quad (34) \)

Note that we have the probability of error as a decreasing function of \( N \), since the \( Q \) function is a decreasing function. We convert an upper bound on the probability of error into a lower bound on \( N \).

\[ P(\text{error}) \leq 0.01 \]

\[ Q\left(\sqrt{\frac{N}{\sigma}}\right) \leq 0.01 \]

\[ \sqrt{\frac{N}{\sigma}} \geq Q^{-1}(0.01) \quad (35) \]

\[ \frac{N}{\sigma} \geq Q^{-1}(0.01)^2 \sigma^2 \]

\[ N \geq (2.3263)^2 \sigma^2 \]

\[ N \geq 5.4119\sigma^2 \]

**Problem VII.9:** In the lecture on Monday - March 5 - the Professor made an incorrect statement. What was this statement? Why is it incorrect, and what statement did he mean to make - i.e. how do you change the statement to be correct?

The professor incorrectly stated that the distribution of a binomial random variable with \( n \) trials and probability of success \( \frac{\lambda}{n} \) approaches the distribution of a Gaussian random variable with mean \( \lambda \) and variance \( \lambda \).

The correct statement is that as \( n \) increases to infinity the distribution of a binomial random variable approaches the distribution of a Poisson random variable with parameter \( \lambda \). To see this, let \( X \) be distributed binomially with parameters \( n \) and \( \frac{\lambda}{n} \).
\[
\lim_{n \to \infty} P(X = k) = \lim_{n \to \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k (1 - \frac{\lambda}{n})^{n-k} \\
= \frac{\lambda^k}{k!} \lim_{n \to \infty} \frac{n!}{(n-k)!n^k} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-k} \\
= \frac{\lambda^k}{k!} \lim_{n \to \infty} \frac{n(n-1) \ldots (n-k+1)}{n^k} \lim_{n \to \infty} (1 - \frac{\lambda}{n})^n \lim_{n \to \infty} (1 - \frac{\lambda}{n})^{-k} \\
= \frac{\lambda^k}{k!} e^{-\lambda}
\]  

We can distribute the limit operations since all the individual limits exist. Note that this convergence is pointwise over the values that \(X\) can take on - as opposed to uniform over the values that \(X\) can take on. The last limit resulting in an exponential is should be familiar to you from highschool calculus. :-) 

In order to get a Gaussian random variable in the limit we need to look at the limit of a binomial random variable whose probability of success does not decrease as \(n\) increases. Suppose \(X\) is distributed binomially with probability of success \(p\) and with \(n\) trials. Let \(Y = \frac{X - np}{\sqrt{n}}\). Then by the central limit theorem as \(n\) increases to infinity the distribution of \(Y\) approaches the distribution of a standard Gaussian random variable. This is because \(X\) is nothing other than the sum of i.i.d. random variables with finite mean and variance - i.e. Bernouli random variables with mean \(p\) and variance \(p(1-p)\).