This problem set essentially reviews notions of conditional expectation, conditional distribution, and Jointly Gaussian random variables. Not all exercises are to be turned in. Only those with the sign ★ are due on Thursday, September 21 at the beginning of the class. Although the remaining exercises are not graded, you are encouraged to go through them. We will discuss some of the exercises during discussion sections. Please feel free to point out errors and notions that need to be clarified.

**Conditional Expectation and distribution**

**Exercise 2.1. ★**

A lost tourist arrives at a point with 3 roads. The first road brings him back to the same point after 1 hours of walk. The second road brings him back to the same point after 6 hours of travel. The last road leads to the city after 2 hours of walk. There are no signs on the roads.

Assuming that the tourist chooses a road equally likely at all times. What is the mean time until the tourist arrives to the city.

**Solution:**

This exercise uses conditional expectation.

Let $T$ the time it takes to reach the city, and $D_i$ the event the tourist chooses door $i$ for $i = 1, 2, 3$. To compute $E[T]$ we will condition on the door scalper picks at the first time:

$$
$$

$$
= \frac{1}{3}(E[T|D_1] + E[T|D_2] + E[T|D_3])
$$

$$
= \frac{1}{3} (1 + E[T] + 6 + E[T] + 2) \tag{2.1}
$$

where in 2.1 we use the fact that after each return to the point, the tourist does not keep memory of its past chooses.

Rearranging the terms in the last equation, we obtain

$$
E[T] = 9 \tag{2.2}
$$

**Exercise 2.2.**
Exercise 2.3. ★
On the day before the exam, each student entering the GSI’s office will ask a question that will come out for the exam with probability $p$. The number of student going to office hours that day is Poisson distributed with mean $\lambda$.

What is the probability that the GSI does not have to answer an exam question?

Solution:
The key idea in this exercise is again conditioning.

Let $N$ be the number of students who enter the GSI’s office on that day (maybe next Monday! 🌐), and let $A$ be the event that the GSI does not have to answer an exam question.

$$P(A) = \sum_n P(A|N = n)P(N = n)$$

$$= \sum_n (1 - p)^n \frac{\lambda^n}{n!} \exp(-\lambda)$$

$$= \exp(-\lambda) \sum_n \frac{(1 - p)\lambda^n}{n!}$$

$$= \exp(-\lambda) \exp((1 - p)\lambda)$$

$$= \exp(-\lambda p)$$

Exercise 2.4. ★
$X$, $Y$, and $Z$ are jointly defined random variables.

1. Show that $E[X] = E[E[X|Y]]$.


Solution:

1. $E[X] = E[E[X|Y]]$
   First note $E[X|Y]$ is a function of $Y$

   $$E[E[X|Y]] = \int_y E[X|y]f_Y(y)dy$$

   $$= \int_y \left( \int_x xf_X(x|y)dx \right) f_Y(y)dy$$

   "all integrals are finite" $= \int_x \left( \int_y f_X(x|y)f_Y(y)dy \right) dx$

   $= \int_x \left( \int_y f_{X,Y}(x, y)dy \right) dx$

   $= \int_x xf_X(x)dx$

   $= E[X]$
2. \( E[Y|Z] = E[E[Y|X, Z]|Z] \)

First note that

\[
E[Y|X, Z] = \int yf_{Y|X,Z}(y|x, z)dy = h(X, Z)
\]

is a function of \((X, Z)\).

Now

\[
E[E[Y|X, Z]|Z] = E[h(X, Y)|Z] = \int x h(x, Z)f_{X|Z}(x|Z)dx
\]

"all integral finite" = \[
\int y \left( \int x f_{Y|X,Z}(y|x, Z)f_{X|Z}(x|Z)dx \right)dy
\]

\[
= \int y \left( \int x f_{Y,X|Z}(y, x)dx \right)dy
\]

\[
= \int y f_{Y|Z}(y|Z)dy
\]

\[
= E[Y|Z]
\]

Exercise 2.5.

Exercise 2.6.

Exercise 2.7. ★

A scalper is considering buying tickets for the next cal football game. The price of the tickets is $20, and the scalper will sell them at $40. However, if he cant sell them at $40, he wont sell them at all. Given that the demand for tickets is a binomial random variable with parameters \(n = 10\) and \(p = 1/2\), how many tickets should he buy in order to maximize his expected profit?

Solution:

Let \(N\) be the number of tickets that the scalper buys, \(D\) be the demand. If \(S\) is the total number of tickets sold, we have \(S = \min(N, D)\) and the scalper’s profit \(R(N)\) is:

\[
R(N) = 40S - 20N
\]

Thus \(E[R(N)] = E[40S - 20N] = 40E[S] - 20N\). Note that here we have assumed that \(N\) is know (we will vary it later). So the only random term is \(S\) (through \(D\)).

Let’s first compute \(E[S]\). We consider that \(N \leq 10\) because it is obvious that buying more
than the maximum demand cannot be optimal. So we have:

\[
E[S] = E[D|D \leq N]P(D \leq N) + E[b|D > N]P(D > N)
\]

\[
= \sum_{i=0}^{N} \binom{10}{i} \left( \frac{1}{2} \right)^{10} + N \sum_{i=N+1}^{10} \binom{10}{i} \left( \frac{1}{2} \right)^{10}
\]

\[
= \left( \frac{1}{2} \right)^{10} \left( \sum_{i=0}^{N} \binom{10}{i} + N \sum_{i=N+1}^{10} \binom{10}{i} \right)
\]

Thus

\[
r(N) = E[R(N)] = 40 \left( \frac{1}{2} \right)^{10} \left( \sum_{i=0}^{N} \binom{10}{i} + N \sum_{i=N+1}^{10} \binom{10}{i} \right) - 20N
\]

Intuitively, \( r(N) \) increases in \( N \) and then decreases. Thus, to find when \( r(N) \) is maximal, we find the first value of \( N \) for which \( r(N + 1) - r(N) \leq 0 \).

We will plot \( r(N) \) and find \( N \) in the next version of this handout.

Exercise 2.8.

2.1 Jointly Gaussian Random Variables

Exercise 2.9.

Exercise 2.10.

Exercise 2.11. ★

John Game uses his high-speed modem to play network games over the internet. The modem transmits zeros and ones by sending signals \(-1\) and \(+1\), respectively. We assume that any given bit has probability \( p \) of being a zero. The telephone line introduces additive zero-mean Gaussian (normal) noise with variance \( \sigma^2 \) (so, the receiver at the other end receives a signal which is the sum of the transmitted signal and the channel noise). The value of the noise is assumed to be independent of the encoded signal value.

1. Let \( a \) be a constant between \(-1\) and \( 1 \). The receiver at the other end decides that the signal \(-1\) (respectively, \(+1\)) was transmitted if the value it receives is less (respectively, more) than \( a \). Find a formula for the probability of making an error.

2. Find a numerical answer for the question of part (a) assuming that \( p = 2/5, a = 1/2 \) and \( \sigma^2 = 1/4 \).

Solution

This exercise considered a detection problem. You will learn more about detection in the next homeworks.
An error occurs whenever $-1$ was transmitted and the received signal is more than $a$ so $-1 + N > a$, or whenever $a+1$ was sent and the received signal is less than $a$ or $1 + N < a$ where $N$ is the noise. The probability of error thus given by:

$$P(error) = P(error|send - 1)P(send - 1) + P(error|send - 1)P(send - 1)$$
$$= P(-1 + N > a)p + P(1 + N < a)(1 - p)$$
$$= P(N > a + 1)p + P(N < a - 1)(1 - p)$$
$$= P(N > a + 1)p + (1 - P(N ≤ a - 1))(1 - p)$$
$$= P\left(\frac{N}{\sigma} > \frac{a + 1}{\sigma}\right) p + (1 - P\left(\frac{N}{\sigma} ≥ \frac{a - 1}{\sigma}\right)) (1 - p)$$

$$= Q\left(\frac{a + 1}{\sigma}\right) p + (1 - Q\left(\frac{a - 1}{\sigma}\right)) (1 - p)$$

where

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

is the complementary cumulative distribution function (CCDF) of a standard normal random variable (that you will study in the next homework!)

There is no closed form for $Q(\cdot)$, but it is a standard function that you will be using quite often.

We can plug in the values and look at a table to find the error.
Will give the values in next version.

Exercise 2.12. ★

The Cauchy distribution has a form

$$f_X(x) = \frac{1}{\pi(1 + x^2)}, \quad -\infty < x < \infty$$

1. Show that this arises from the ratio $X_1/X_2$ where $X_1$ and $X_2$ are independent zero-mean gaussian random variable with variance $\sigma^2$.

2. Show that the moments of $X$ do not exist.

Solution:
To show (1), we will compute the pdf of the random variable $Z = \frac{X_1}{X_2}$ and observe that it
has a Cauchy distribution. We have:

\[ F_Z(z) = P(Z \leq z) = P \left( \frac{X_1}{X_2} \leq z \right) \]
\[ = \int_{x_2} P \left( \frac{X_1}{x_2} \leq z | x_2 \right) f_{X_2}(x_2) dx_2 \]
\[ = \int_{x_2} P(X_1 < zx_2) f_{X_2}(x_2) dx_2 \quad \text{Using independence} \]
\[ = \int_{x_2} F_{X_1}(zx_2) f_{X_2}(x_2) dx_2 \]

Now we can compute the pdf of \( Z \) by taking the derivative of the last expression with respect to \( z \).

\[ f_Z(z) = \frac{\partial}{\partial z} \int_{x_2} F_{X_1}(zx_2) f_{X_2}(x_2) dx_2 \]

Since the integral is finite (and everything is nice! Why?) and \( z \) is not involved in the integration, we can safely interchange the order of the derivative and the integral to obtain:

\[ f_Z(z) = \frac{\partial}{\partial z} \int_{x_2} x_2 f_{X_1}(zx_2) f_{X_2}(x_2) dx_2 \]
\[ = \frac{1}{2\pi \sigma^2} \int_{x_2} x_2 e^{-\frac{(zx_2)^2}{\sigma^2}} e^{-\frac{(x_2)^2}{\sigma^2}} dx_2 \]
\[ = \frac{1}{2\pi \sigma^2} \int_{x_2} x_2 e^{-\frac{1+zx_2^2}{2\sigma^2}} dx_2 \]
\[ = \frac{1}{\pi \sigma^2} \left[ -\frac{\sigma^2}{1+z^2} e^{-\frac{1+zx_2^2}{2\sigma^2}} \right]_0^\infty \]
\[ = \frac{1}{\pi} \frac{1}{1+z^2} \]

which is the pdf of a Cauchy random variable.

To show that the moments do not exist, we will show that \( \int \frac{x^k}{1+x^2} dx \) does not exist for any \( k \geq 1 \).

First observe that

\[ \frac{x^k}{1+x^2} \geq \frac{x}{1+x^2} \quad \text{for } k > 0 \text{ and } x \geq 1 \]

But

\[ \int_1^\infty \frac{x}{1+x^2} dx = \lim_{T \to \infty} \frac{1}{2} \log(1+T^2) = \infty \]
Now
\[
\int_{-\infty}^{\infty} \frac{|x|^k}{1+x^2} \, dx \geq \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} \, dx, \forall k
\]
\[
= C^* + 2 \int_{1}^{\infty} \frac{x}{1+x^2} \, dx
\]
\[
= \infty
\]
Thus the functions $\frac{x^k}{1+x^2}$ are not integrable for all $k \geq 1$. Hence the moments do not exist.

Exercise 2.13.
Exercise 2.14. ★

Let
\[
\begin{align*}
Z_1 &= Y_1 + Y_2 \\
Z_2 &= Y_2 \\
Z_3 &= Y_1 - Y_2
\end{align*}
\]
for some independent standard normal random variables $Y_1$ and $Y_2$.

1. Is $Z = (Z_1, Z_2, Z_3)^T$ JG?
2. If yes compute its correlation matrix. Does it have an inverse?
3. Now you are given a JG random vector $W$ with correlation matrix:
\[
K_W = \begin{pmatrix}
2 & 0 & 1 \\
0 & 6 & -3 \\
1 & -3 & 2
\end{pmatrix}
\]
Find matrix $M$ such that $W = MV$ for JG $V$ with $\det K_V \neq 0$.
Hint: Use eigenvalue decomposition

Solution:
Since the $Z_i$‘s are linear combinations of independent standard normal random variables, the vector $Z$ is JG.
Its correlation matrix is:
\[
K_Z = \begin{pmatrix}
2 & 1 & 0 \\
1 & 1 & -1 \\
0 & -1 & 2
\end{pmatrix}
\]
Since determinant is zero, $K_Z$ does not admit an inverse.
We will solve (3) for a general $K_W \in \mathbb{R}^{n \times n}$ and let you verify your result using matlab or any other software.
The idea of this exercise is to write any JG random vector $W$, with $|K_W| = 0$, as a function of another JG random $Z$ that has non-singular correlation matrix. *If we succeed to do so* we can use $Z$ to compute things like conditional mean and pdf given $Y$, since we can always go
from \( Z \) to \( W \).
Using E.V.D., we can write \( K_W = Q\Lambda Q^T \) for some orthogonal \( Q \), and diagonal matrix
\[
\Lambda = \begin{pmatrix} 
\Lambda_m & 0 \\
0 & 0 
\end{pmatrix}
\]
for some \( m < n, \Lambda_m \in \mathbb{R}^{m \times m}, |\Lambda_m| > 0. \)
Set
\[
M = Q \begin{pmatrix}
\Lambda_m^{\frac{1}{2}} \\
0
\end{pmatrix}
\]
Now we will solve \( W = MZ \) for \( Z \in \mathbb{R}^m. \)
\[
W = Q \begin{pmatrix}
\Lambda_m^{\frac{1}{2}} \\
0
\end{pmatrix} Z 
\]
\[
Q^T W = \begin{pmatrix}
\Lambda_m^{\frac{1}{2}} \\
0
\end{pmatrix} Z 
\]
\[
(\Lambda_m^{-\frac{1}{2}} 0)Q^T W = (\Lambda_m^{-\frac{1}{2}} 0) \begin{pmatrix}
\Lambda_m^{\frac{1}{2}} \\
0
\end{pmatrix} Z = I_m Z
\]
\[
(\Lambda_m^{-\frac{1}{2}} 0)Q^T W = Z
\]
By defining \( B := (\Lambda_m^{-\frac{1}{2}} 0)Q^T \), we have \( Z = BW \) and \( W = MZ \). Also
\[
K_Z = E[BB^T] = (\Lambda_m^{-\frac{1}{2}} 0)Q^T \Lambda Q^T Q \begin{pmatrix}
\Lambda_m^{-\frac{1}{2}} \\
0
\end{pmatrix} = I_m
\]
Thus \(|K_Z| \neq 0\), and also \( Z \) is JG.

**Exercise 2.15. ★**

Let \( X_1 \) and \( X_2 \) be two independent standard normal random variables and let
\[
\begin{cases}
Y_1 = X_1 + X_2 \\
Y_2 = X_1 - X_2 \\
Y_3 = 3X_1 + X_2 \\
\end{cases}
\]
Compute the conditional pdf \( f_{Y_1|Z}(y_1|y_2,y_3) \) of \( Y_1 \) given \( (Y_2,Y_3) \) and compute its mean.

**Hint:** Argue that \( Y_1, Y_2, \) and \( Y_3 \) are jointly gaussian and show \( (Y_2,Y_3)^T = AZ \) and \((Y_1,Z)\) is JG, for some matrix \( A \), and \( Z \) with \( \det(K_Z) \neq 0. \) Calculate \( f_{Y_1|Z}(\cdot|\cdot), K_{Y_1,Z}, \) and \( K_Z^{-1} \)

**Remark:**
In this exercise I did not choose good coefficients to make the problem interesting. As I promised in discussion sections, I will present another example that illustrates better the point I wanted you to get.
Exercise 2.16. ★
Gallager’s note: Exercise 2.10

Solution:
(a) Note that there is a one-to-one mapping between \( Y \) and \( Y^3 \) meaning that given \( Y^3 \) we can uniquely deduce \( Y \). Thus the distribution of \( X \) remains the same given that we have observed \( Y \) or \( Y^3 \). We can now use formula (2.24) of the reader \( f_{X|Y}(x|y) \) (we replace \( y \) by \( \sqrt{y} \) because we observe \( Y^3 \)).

\[
f_{X|Y^3}(x|v) = \frac{1}{\sigma_x\sqrt{2\pi(1-\rho)^2}} \exp\left(-\frac{(x - \rho(\sigma_x/\sigma_y)\sqrt{v})^2}{2\sigma_x^2(1-\rho)^2}\right)
\]

(b) This part is a little harder because we no longer have the one-to-one mapping. However we can apply the technique in part (a) if we first condition on the sign of \( Y \):

\[
f_{X|Y^2}(x|v) = f_{X|Y^2}(x|v, Y > 0)P(Y > 0|Y^2 = v) + f_{X|Y^2}(x|v, Y \leq 0)P(Y \leq 0|Y^2 = v)
\]

The event \( \{Y > 0\} \) is independent of \( Y^2 \) (convince yourself!), so \( P(Y > 0|Y^2 = v) = P(Y > 0) = 1/2 \). For \( v > 0 \), the events \( \{Y^2 = v, Y > 0\} \) and \( \{Y = \sqrt{v}\} \) are the same, so

\[
f_{X|Y^2}(x|v, Y > 0) = \frac{1}{\sigma_x\sqrt{2\pi(1-\rho)^2}} \exp\left(-\frac{(x - \rho(\sigma_x/\sigma_y)\sqrt{v})^2}{2\sigma_x^2(1-\rho)^2}\right)
\]

Similarly, for \( v \geq 0 \) the events \( \{Y^2 = v, Y \leq 0\} \) and \( \{Y = -\sqrt{v}\} \) are the same, so

\[
f_{X|Y^2}(x|v, Y > 0) = \frac{1}{\sigma_x\sqrt{2\pi(1-\rho)^2}} \exp\left(-\frac{(x + \rho(\sigma_x/\sigma_y)\sqrt{v})^2}{2\sigma_x^2(1-\rho)^2}\right)
\]

Thus

\[
f_{X|Y^2}(x|v) = \frac{1}{2\sigma_x\sqrt{2\pi(1-\rho)^2}} \left( \exp\left(-\frac{(x - \rho(\sigma_x/\sigma_y)\sqrt{v})^2}{2\sigma_x^2(1-\rho)^2}\right) + \exp\left(-\frac{(x + \rho(\sigma_x/\sigma_y)\sqrt{v})^2}{2\sigma_x^2(1-\rho)^2}\right) \right)
\]

Exercise 2.17. ★
Gallager’s note: Exercise 2.13

(a) For \( Y = AW \),

\[
\]

Thus in this case writing \( K = [E[Wy^T]] = A^T \), we see that \( A = K^T \) achieves the desired cross-covariance.

(b) We proceed as in part (a) and compute

\[
E[WZ^T] = E[WW^T] B^T = K Z B^T
\]
Thus \( B = K^T K_Z^{-1} \) will achieve the desired cross-correlation.

(c) From part (b), we know that we are looking for \( B \) such that \( K_Z B^T = K \). Assuming that \( B^T = [B_1 | \ldots | B_m] \) and \( K = [K_1 | \ldots | K_m] \), let us rewrite \( K_Z B^T = K \) in the following form

\[
K_Z[B_1 | \ldots | B_m] = [K_1 | \ldots | K_m]
\]

This can be decomposed into a set of linear systems of equations \( K_Z B_i = K_i, i = 1, \ldots, m \).

Now observe that \( K_Z = Q \Lambda Q^T \) where \( Q \) is orthogonal and \( \Lambda \) has the form

\[
\Lambda = \begin{pmatrix}
\lambda_1 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \lambda_r & 0 & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0
\end{pmatrix}
\]

\( K_Z B_i = K_i \) can be written as \( \Lambda Q^T B_i = Q^T K_i \) or in a more detailed form:

\[
\begin{cases}
\lambda_1 Q_1^T B_i = Q_1^T K_i \\
\vdots \\
\lambda_r Q_r^T B_i = Q_r^T K_i \\
0 = Q_{r+1}^T K_i \\
\vdots \\
0 = Q_n^T K_i
\end{cases}
\]

This tells us that \( K_i, i = 1, \ldots, m \) is orthogonal to \( Q_j, j = r + 1, \ldots, n \). In other words, each column of \( K \) is orthogonal to all the eigenvectors that have a zero eigenvalue. This is equivalent to saying that each column of \( K \) belongs to \( V = \text{span}(Q_1, \ldots, Q_r) \).

In conclusion, for \( B \) to satisfy the desired cross-correlation matrix, each column of \( K \) must belong to the space spanned by the eigenvectors that correspond to the non-zero eigenvalues of \( K_Z \).

**Complement**

2.1.1 Proof of the sufficient statistics property

In class we have stated that if \( g(Y) \) is a sufficient statistic for observation \( Y \), then \( X \) is independent to \( Y \) given \( g(Y) \). We give a proof of this fact below.

The proof is given for \( X \) discrete but it can easily be generalized to the continuous case.

Consider a detection problem where the observation \( Y \) is used to decide on the value of \( X \). We assume that \( X \) lies in a discrete set \( I = 1, \ldots, n \) and its prior probability is given by \( P(X = x) = \pi(x) \). The conditional distribution of the observation \( Y \) given \( X \) is given by
\( f_{Y|X}(y|x) \).

**Fact:**
If \( g(Y) \) is a sufficient statistic for \( Y \), then \( X \) is independent to \( Y \) given \( g(Y) \).

**Proof:** To prove the fact, we will show that \( f_{Y|(X,g(Y))}(y|x, g(y)) \) does not depend on \( X \).

First note that since \( g(Y) \) is a sufficient statistic, the conditional distribution of \( Y \) given \( X \) can be written in the form:

\[
f_{Y|X}(y|x) = F(g(y), x)h(y)
\]

where \( F(.) \) depends on \( Y \) only through \( g(.) \) and \( h(.) \) is only a function of \( y \).

Now we have:

\[
f_{Y|(X,g(Y))}(y|x, g(y) = g) = \frac{f_{Y,X,g(Y)}(y, x, g(y) = g)}{f_{X,g(Y)}(x, g(y) = g)}
\]

\[
= \frac{\pi(x) f_{Y|X}(y|x)1_{[g(y) = g]}}{\pi(x) \sum_{y'} f_{Y|X}(y'|x)1_{[g(y') = g]}}
\]

\[
= \frac{F(g(y), x)h(y)1_{[g(y) = g]}}{\sum_{y'} F(g(y'), x)h(y')1_{[g(y') = g]}}
\]

\[
= \frac{h(y)1_{[g(y) = g]}}{\sum_{y'} h(y')1_{[g(y') = g]}}
\]

where we go from 2.4 to 2.5 by observing that \( f_{Y,X,g(Y)}(y, x, g(y) = g) \) is 0 for any \( y \) such that \( g(y) \neq g \) and is otherwise equal to \( f_{Y|X}(y|x) \). In 2.6, we have used 2.3 and canceled out the \( \pi(x) \)'s. The term \( F(.) \) in the denominator is not involved in the summation, so we take it out in 2.7 and it cancels out the corresponding term in the numerator and we see that \( f_{Y|(X,g(Y))}(y|x) \) does not depend on \( x \). \( \square \)

### 2.1.2 Remake of Exercise 2.15

You have all noticed that exercise 2.15 was too obvious. In fact there was an error in the choice of the coefficients and the exercise did not illustrate what it was supposed to illustrate. We give here an example that illustrates better the notion.

**Example:**
Let \((X, Y_1, Y_2)^T\) is a zero-mean jointly gaussian random vector with correlation matrix

\[
K = \begin{pmatrix}
5 & 3 & 1 \\
3 & 9 & 3 \\
1 & 3 & 1
\end{pmatrix}
\]

Find \( E[X|Y] \).

In this example we see that \(|K_Y| = 0\) and we cannot apply the formula given in class

\[
E[X|Y] = K_{XY}K^{-1}_Y Y
\]

2-11
The hint in exercise (2.15) was to find $Z$ such that $Y = AZ$, $(X, Z)$ is JG, and $|K_Z| \neq 0$. In exercise (2.14) part (3), we have seen one way to compute $A$ and $Z$. Here we will give a less formal procedure.

First notice that the second row of $K_Y$ is a multiple of the first row. Let’s study the random variable $T = Y_1 - 3Y_2$.

$$E[T] = E[Y_1] - 3E[Y_2] = 0$$

$$Var(T) = E[(Y_1 - 3Y_2)^2] = E[Y_1^2] + 9E[Y_2^2] - 6E[Y_1Y_2] = 9 + 9 - 3 \times 6 = 0$$

$T$ has zero mean and zero variance thus it is deterministic and equal to 0. So we have that $Y_1 = 3Y_3$. Hence $Y_1$ carries all the information in $Y = (Y_1, Y_2)$ (or is a sufficient statistic). Thus

$$E[X|Y] = E[X|Y_1]$$

It is not hard to convince yourself that $(X, Y_1)$ is JG and $K_{Y_1} = \sigma_{Y_1}^2 \neq 0$. Thus

$$E[X|Y_1] = \frac{\sigma_{XY_1}}{\sigma_{Y_1}^2}Y_1$$

$$= \frac{1}{3}Y_1$$