WEEK 4

4. Moments of Random Variables (mean, variance, covariance)

In this part of the course we are going to look at the moments of the random variables. As we know, if X is a random variable then any function of a random variable is also a random variable. That is, $g(X)$ is also a random variable.

On the other hand, we categorized the random variables as discrete and continuous. If the range of the random variable is a countable subset of the real line we said that the random variable is discrete and continuous otherwise. That is a random variable X could be either discrete or continuous and therefore, the transformed random variable $g(X)$ will be either discrete or continuous.

Note: If X is a continuous random variable, $g(X)$ could be either continuous or discrete and similarly, if X is discrete then $g(X)$ will be either discrete or continuous. But in our course framework, if X is discrete then $g(X)$ will be discrete and if X is a continuous then $g(X)$ will be a continuous random variable.

Definition: Let X be a random variable with a probability (or probability density) function $f(x)$. The expected value of the random variable $g(X)$ is

$$E(g(X)) = \begin{cases} \sum\limits_{x \in D_X} g(x) f(x) & , \quad \sum\limits_{x \in D_X} |g(x)| f(x) < \infty \\ \int\limits_{x \in D_X} g(x) f(x) dx & , \quad \int\limits_{x \in D_X} |g(x)| f(x) dx < \infty. \end{cases}$$

Definition: Let X be a random variable with a probability (or probability density) function $f(x)$. Assume that $E(g(X))$ exists. Then

a) If $g(x) = x$ then $E(X)$ is called the expected value (or mean) of the random variable X and denoted by μ.

b) If $g(x) = x^k$ then $E(X^k)$ is called the k^{th} moment of the random variable X and denoted by μ_k.

c) The number $\mu_2 - \mu_1^2$ is called the variance of the random variable X and denoted by either σ^2 or $Var(X)$. That is, $Var(X) = E(X^2) - (E(X))^2$.

d) The positive square root of the variance is called the standard deviation of the random variable X and denoted by σ. That is, $\sigma = +\sqrt{Var(X)}$.

e) The function $E(e^{tX})$ is called the moment generating function of the random variable X and denoted by $M_X(t)$. That is, $M_X(t) = E(e^{tX})$.

f) The median of a random variable X is a number m such that $P(X \geq m) \geq 1/2$ and $P(X \leq m) \geq 1/2$.

g) The number γ ,

$$\gamma = \frac{E(X-\mu)^3}{\sigma^3}$$

is called the skewness of the random variable X .

h) g) The number κ ,

$$\kappa = \frac{E(X-\mu)^4}{\sigma^4} - 3$$

is called the kurtosis of the random variable X .

<u>Example</u>: Let X be a random variable with a probability function

$$P(X = x) = p^x q^{1-x} \text{ , } x = 0{,}1 \text{ and } q = 1 - p.$$

First of all, for any integer $k \in \mathbb{N}$ we have

$$E(X^k) = \sum_{x=0}^{1} x^k P(X = x) = \sum_{x=0}^{1} x^k p^x q^{1-x} = 0^k P(X = 0) + 1^k P(X = 1) = P(X = 1) = p \text{ .}$$

That is any moment of the random variable p. Thus, for any integer k we have $E(X^k) = \mu_k = p.$

Therefore,

$$E(X) = \mu = p \text{ , } E(X^2) = \mu_2 = p$$

which implies that

$$Var(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p) = pq \text{ .}$$

The standard deviation is $\sigma = +\sqrt{Var(X)} = +\sqrt{pq}$. The moment generating function of the random variable

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{1} e^{tx} P(X = x) = e^{0t} P(X = 0) + e^t P(X = 1) = q + pe^t \text{ .}$$

That is, $M_X(t) = q + pe^t$.

Note that in general we have

$$\frac{d^k}{dt^k}M_X(t)\Big|_{t=0} = \frac{d^k}{dt^k}E\left(e^{tX}\right)\Big|_{t=0} = E\left(X^k e^{tX}\right)\Big|_{t=0} = E(X^k) = \mu_k .$$

In our case, as we have calculated above

$$\frac{d^k}{dt^k}M_X(t)\Big|_{t=0} = \frac{d^k}{dt^k}\left(q+pe^t\right)\Big|_{t=0} = pe^t\Big|_{t=0} = p = E(X^k) = \mu_k .$$

The skewness: In order to calculate the skewness of the random variable, first we calculate,

$$E(X-\mu)^3 = \sum_{x=0}^{1}(x-\mu)^3 P(X=x) = \sum_{x=0}^{1}(x-p)^2\, p^x q^{1-x} = (0-p)^3 P(X=0)+(1-p)^3 P(X=1)$$
$$= (-p)^3 q + q^3 p = pq(q^2-p^2)$$

and therefore, the skewness is

$$\gamma = \frac{E(X-\mu)^3}{\sigma^3} = \frac{pq(q^2-p^2)}{(pq)^{3/2}} = \frac{(q^2-p^2)}{(pq)^{1/2}} = \frac{(q-p)(q+p)}{\sigma} = \frac{(q-p)}{\sigma} .$$

Notes: : Let X be a random variable with a probability (or probability density) function $f(x)$ and $a,b\in\mathbb{R}$ Then

a) $E(a+bX) = a+bE(X)$ b) $Var(a+bX) = b^2E(X)$

Proof: The results are true for both discrete and continuous case. In the following, we will have the proof for discrete case. For continuous case you only need to change the sums with integrals

a) Let $a,b\in\mathbb{R}$ then (. The range of the random variable is D_X)

$$E(a+bX) = \sum_{x\in D_X}(a+bx)P(X=x) = a\sum_{x\in D_X}P(X=x) + b\sum_{x\in D_X}P(X=x) = a+bE(X)$$

b) The variance of $a+bX$ can be calculated directly as

$$Var(a+bX) = E(a+bX)^2 - (E(a+bX))^2 = E(a^2+2abX+b^2X^2) - \left[a+bE(X)\right]^2$$
$$= E(a^2+2abX+b^2X^2) - [a^2+2abE(X)+b^2(E(X))^2]$$
$$= \cancel{a^2} + \cancel{2abE(X)} + b^2E(X^2) - \cancel{a^2} - \cancel{2abE(X)} - b^2(E(X))^2 = b^2E(X^2) - b^2(E(X))^2$$
$$= b^2[E(X^2)-(E(X))^2] = b^2Var(X)$$

which completes the proof.

Note that,

$$E(X-\mu)^2 = E(X^2-2\mu X+\mu^2) = E(X^2)-2\mu E(X)+\mu^2 = E(X^2)-2\mu^2+\mu^2$$
$$= E(X^2)-\mu^2 = E(X^2)-(E(X))^2 = Var(X).$$

Here, $\mu = E(X)$. That is, the variance of a random variable can be calculated as either $E(X^2)-(E(X))^2$ or $E(X-\mu)^2$.

Example: (continuous case)

Let X be a continuous random variable with the following probability density function

$$f(x)=\begin{cases}\dfrac{1}{\beta}e^{-x/\beta} & , \quad x>0 \\ 0 & , \quad elsewhere.\end{cases}$$

Remember the properties of Gamma function from your calculus class as

$$\Gamma(\alpha)\beta^{\alpha} = \int_{x=0}^{\infty} x^{\alpha-1}e^{-x/\beta}dx\,,\ \Gamma(\alpha+1)=\alpha\,\Gamma(\alpha) \ \text{ and } \ \Gamma(n+1)=n!\ .$$

Now, we can calculate any moment of the random variable as

$$E(X^k)=\int_{x=0}^{\infty} x^k f(x)dx=\frac{1}{\beta}\int_{x=0}^{\infty} x^k e^{-x/\beta}dx=\frac{1}{\beta}\int_{x=0}^{\infty} x^{(k+1)-1}e^{-x/\beta}dx=\frac{\Gamma(k+1)\beta^{k+1}}{\beta}=\Gamma(k+1)\beta^k\ .$$

Thus, we can calculate first two moments and the variance as

for $k=1$ $E(X)=\Gamma(1+1)\beta=\Gamma(2)\beta=\beta$

for $k=2$ $E(X^2)=\Gamma(2+1)\beta^2=\Gamma(3)\beta^2=2\beta^2$

and therefore, the variance is

$$Var(X)=E(X^2)-(E(X))^2=2\beta^2-\beta^2=\beta^2\ .$$

Multivariate case:

Let X and Y be two random variables with joint probability (or probability density) function $f(x,y)$. We know that if X and Y are two random variables then any function of these random variables is also a random variable. That is $g(X,Y)$ is also a random variable. From the joint probability (or probability density) function we can calculate the marginal probability (or probability density) functions of X and Y as (for discrete case you only need to change the integral with summations)

$$f_X(x)=\int_{y\in D_Y} f(x,y)\,dy \quad \text{and} \quad f_Y(y)=\int_{x\in D_X} f(x,y)\,dx$$

Using these marginal probability (or probability density) functions we can calculate $E(X)$, $E(Y)$, $Var(X)$ and $Var(Y)$ as we have discussed above. Let g be any function from $\mathbb{R}^2$ to $\mathbb{R}$ (that is, $g:\mathbb{R}^2\to\mathbb{R}$). Then the expected value of $g(X,Y)$ is given by

$$E(g(X,Y))=\begin{cases}\displaystyle\sum_{x\in D_X}\sum_{y\in D_Y} g(x,y)P(X=x,Y=y) & , \quad \displaystyle\sum_{x\in D_X}\sum_{y\in D_Y}|g(x,y)|P(X=x,Y=y)<\infty \\ \displaystyle\int_{x\in D_X}\int_{y\in D_Y} g(x,y)f(x,y)\,dy\,dx & , \quad \displaystyle\int_{x\in D_X}\int_{y\in D_Y}|g(x,y)f(x,y)|\,dy\,dx<\infty.\end{cases}$$

If we set $g(x,y) = x\,y$, the expected value of XY can be calculated by either

$$E(XY) = \sum_{x\in D_X} \sum_{y\in D_Y} xy\,P(X=x, Y=y) \quad \text{or} \quad E(XY) = \int_{x\in D_X} \int_{y\in D_Y} g(x,y)\,f(x,y)\,dy\,dx\,.$$

Now we have $E(X)$, $E(Y)$ and $E(XY)$.

Definition: The difference between $E(XY)$ and $E(X)E(Y)$ is called the covariance between X and Y denoted by $Cov(X,Y)$. That is $Cov(X,Y) = E(XY) - E(X)E(Y)$. The covariance between X and Y can also be calculated as $Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$.

From the marginal probability (or probability density) function we can also calculate the variances. Using these values, we define the correlation between to random variables.

Definition: The ration of $Cov(X,Y)$ to the square root of $Var(X)Var(Y)$ is called the correlation between X and Y denoted by ρ_{XY}. That is,

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}\,.$$

The correlation between two random variable measures some kind of dependency. In general, the correlation is always between -1 and $+1$, that is $-1 \le \rho_{XY} \le 1$ or $0 \le \rho_{XY}^2 \le 1$.

Note: If X and Y are independent random variables then the correlation is zero but the converse is not true in general. The result is true for both discrete and continuous case. The proof will be done for discrete case. For continuous case you and change the summations to integrals.

Proof: Let X and Y be two independent random variables then the joint probability function can be written as a product of the marginal probability functions. That is, $f(x,y) = f_X(x)\,f_Y(y)$ for all x and y in $\mathbb{R}$. Therefore,

$$\begin{aligned} E(XY) &= \sum_{x\in D_X} \sum_{y\in D_Y} x\,y\,f(x,y) = \sum_{x\in D_X} \sum_{y\in D_Y} x\,y\,f_X(x)\,f_Y(y) \\ &= \sum_{x\in D_X} x\,f_X(x) \sum_{y\in D_Y} y\,f_Y(y) = E(X)E(Y) \end{aligned}\,.$$

That is, if X and Y are two independent random variables then $E(XY) = E(X)E(Y)$ which implies that $Cov(X,Y) = E(XY) - E(X)E(Y) = 0$ and therefore $\rho_{XY} = 0$.

Notes:

1. If X and Y are two independent random variables then $E(XY) = E(X)E(Y)$ but the converse is NOT true. That is, having $\rho_{XY} = 0$ does not imply independency. However, in this class if $\rho_{XY} = 0$ you can assume that the random variables are independent.

2. If $\rho_{XY} = 1$ then the random variables are linearly dependent. This means that you can find constants $a, b \in \mathbb{R}$ such that $P(Y = a + bX) = 1$.

Example: Let X and Y be two random variables having the following joint probability function:

$$f(x,y) = \begin{cases} c(x+y) & , \quad x = 0,1,2\ ;\ y = 1,2 \\ 0 & , \quad elsewhere. \end{cases}$$

Let us calculate the correlation between these two random variables. First we need to determine the value c. Note that

$$1 = \sum_{x=0}^{2} \sum_{y=1}^{2} c(x+y) = c\sum_{x=0}^{2} [(x+1)+(x+2)] = c\sum_{x=0}^{2} (2x+3) = c[(0+3)+(2+3)+(4+3)] = 15c$$

which implies that $c = 1/15$. That is the joint probability function of X and Y can be written as

$$f(x,y) = \begin{cases} (x+y)/15 & , \quad x = 0,1,2\ ;\ y = 1,2 \\ 0 & , \quad elsewhere. \end{cases}$$

Now we can find the marginal probability functions by using the following summations:

$$f_X(x) = \frac{1}{15}\sum_{y=1}^{2} (x+y) = \frac{1}{15}[(x+1)+(x+2)] = \frac{2x+3}{15}$$

$$f_Y(y) = \frac{1}{15}\sum_{x=0}^{2} (x+y) = \frac{1}{15}[(0+y)+(1+y)+(2+y)] = \frac{3y+3}{15} = \frac{y+1}{5}.$$

Thus, the marginal probability functions of X and Y are

$$f_X(x) = \begin{cases} (2x+3)/15 & , \quad x = 0,1,2 \\ 0 & , \quad elsewhere \end{cases} \quad \text{and} \quad f_Y(y) = \begin{cases} (y+1)/5 & , \quad y = 1,2 \\ 0 & , \quad elsewhere. \end{cases}$$

From the marginal probability functions we can calculate the mean and variances of X and Y .

$$E(X) = \sum_{x=0}^{2} x\, f_X(x) = \frac{1}{15}\sum_{x=0}^{2} x(2x+3) = \frac{1}{15}[0+5+14] = \frac{19}{15}$$

$$E(X^2) = \sum_{x=0}^{2} x^2\, f_X(x) = \frac{1}{15}\sum_{x=0}^{2} x^2(2x+3) = \frac{1}{15}[0+5+28] = \frac{33}{15} = \frac{11}{5}$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{11}{5} - \frac{(19)^2}{225} = \frac{45(11) - 19(19)}{225} = \frac{134}{225}$$

Similarly,

$$E(Y) = \sum_{y=1}^{2} y\, f_Y(y) = \frac{1}{5}\sum_{y=1}^{2} y(y+1) = \frac{1}{5}[2+6] = \frac{8}{5}$$

$$E(Y^2)=\sum_{y=1}^{2} y^2\ f_Y(y)=\frac{1}{5}\sum_{y=1}^{2} y^2(y+1)\ =\frac{1}{5}[2+12]=\frac{14}{5}$$

$$Var(Y)=E(Y^2)-(E(Y))^2=\frac{14}{5}-\frac{64}{25}=\frac{5(14)-64}{25}=\frac{70-64}{25}=\frac{6}{25}$$

Finally, we need $E(XY)$ which is calculated as

$$E(XY)=\frac{1}{15}\sum_{x=0}^{2}\sum_{y=1}^{2} xy(x+y)=\frac{1}{15}\sum_{x=0}^{2}[x(x+1)+2x(x+2)]=\frac{1}{15}\sum_{x=0}^{2}(3x^2+5x)$$
$$=\frac{1}{15}\Big[[3(1)^2+5(1)]+[3(2)^2+5(2)]\Big]=\frac{1}{15}[8+22]=2.$$

Thus,

$$Cov(X,Y)=E(XY)-E(X)-E(Y)=2-\frac{19}{15}\,\frac{8}{5}=2-\frac{152}{75}=\frac{150-152}{75}=-\frac{2}{75}$$

and the correlation between X and Y is

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}=\frac{-2/75}{\sqrt{(134/225)(6/25)}}=-\frac{2}{75}\frac{\sqrt{(225)(25)}}{\sqrt{(134)(6)}}=-\frac{2}{\sqrt{804}}\cong-0.0705\,.$$

Now, let us calculate the conditional expectation of Y given $X=x$. That is we want to calculate $E(Y\,|\,X=x)$. In order to calculate this conditional expectation, first we need the conditional probability function of Y given $X=x$. This is found to be

$$f_{Y|X=x}(y\,|\,x)=\frac{f(x,y)}{f_X(x)}=\frac{(x+y)/15}{(2x+3)/15}=\frac{(x+y)}{(2x+3)}\ ,\ y=1,2$$

and the conditional expectation of Y given $X=x$ is

$$E(Y\,|\,X=x)=\sum_{y=1}^{2} y f_{Y|X=x}(y\,|\,x)=\sum_{y=1}^{2} y\left(\frac{(x+y)}{(2x+3)}\right)=\frac{x+1}{2x+3}+\frac{2(x+2)}{2x+3}=\frac{3x+5}{2x+3}\,.$$

<u>Example</u>: (continuous case) Let X and Y be two random variables having the following joint probability density unction:

$$f(x,y)=\begin{cases} c(x+y) & ,\quad 0<x<2\,;0<y<2 \\ 0 & ,\quad elsewhere. \end{cases}$$

Let us calculate the correlation between these two random variables. First we need to determine the value c. Note that

$$1=c\int_{x=0}^{2}\int_{y=0}^{2}(x+y)\,dy\,dx=c\int_{x=0}^{2}\Big[xy+y^2/2\Big]_{y=0}^{2}dx=c\int_{x=0}^{2}(2x+2)\,dx=c[x^2+2x]_{x=0}^{2}=8c$$

which imlies that $c=1/8$. That is the joint probability density function is

$$f(x,y)=\begin{cases}(x+y)/8 & , \quad 0<x<2 \ ; \ 0<y<2\\ 0 & , \quad elsewhere \quad .\end{cases}$$

In order to calculate the covariance (or correlation) between these two random variables first we need the marginal probability density functions. Note that

$$\frac{1}{8}\int_{y=0}^{2}(x+y)\,dy=\frac{x+1}{4} \quad \text{and} \quad \frac{1}{8}\int_{x=0}^{2}(x+y)\,dx=\frac{y+1}{4}$$

and thus the marginal probability density functions can be calculayed as

$$f_X(x)=\begin{cases}(x+1)/4 & , \quad 0<x<2\\ 0 & , \quad elsewhere\end{cases} \qquad f_Y(y)=\begin{cases}(y+1)/4 & , \quad 0<y<2\\ 0 & , \quad elsewhere.\end{cases}$$

Note that the random variables X and Y have the same probability density function (that is, they are identically distributed) but they are not independently distributed random variables because

$$f(x,y)=\left(\frac{x+y}{8}\right)\neq\left(\frac{x+1}{4}\right)\left(\frac{y+1}{4}\right)=f_X(x)\,f_Y(y)\,.$$

Let us try to calculate the probabilities: $P(X<1)$, $P(Y>1)$ and $P(X<1|Y>1)$. From the marginal probability density functions the probabilities are calculated as:

$$P(X<1)=\int_{x=0}^{1}f_X(x)\,dx=\frac{1}{4}\int_{x=0}^{1}(x+1)\,dx=\frac{1}{4}\Big[(x^2/2)+x\Big]_{x=0}^{1}=\frac{1}{4}\left[\frac{3}{2}\right]=\frac{3}{8}$$

and

$$P(Y>1)=\int_{y=1}^{2}f_Y(y)\,dx=\frac{1}{4}\int_{y=1}^{2}(y+1)\,dx=\frac{1}{4}\Big[(y^2/2)+y\Big]_{y=1}^{2}=\frac{1}{4}\Big[(2+2)-\big(\tfrac{1}{2}+1\big)\Big]=\frac{5}{8}$$

finally, from the follwing integral

$$P(X<1,Y>1)=\int_{y=1}^{2}\int_{x=0}^{1}f(x,y)\,dx\,dy=\frac{1}{8}\int_{y=1}^{2}\int_{x=0}^{1}(x+y)\,dx\,dy=\frac{1}{8}\int_{y=1}^{2}\left[\frac{x^2}{2}+xy\right]_{x=0}^{1}dy$$

$$=\frac{1}{8}\int_{y=1}^{2}\left[y+\frac{1}{2}\right]dy=\frac{1}{8}\left[\frac{y^2}{2}+\frac{y}{2}\right]_{y=1}^{2}=\frac{1}{8}\left[\left(\frac{4}{2}+\frac{2}{2}\right)-\left(\frac{1}{2}+\frac{1}{2}\right)\right]=\frac{1}{8}\big[(2+1)-1\big]=\frac{2}{8}=\frac{1}{4}$$

we have

$$P(X<1|Y>1)=\frac{P(X<1,Y>1)}{P(Y>1)}=\frac{1/4}{5/8}=\left(\frac{1}{4}\right)\left(\frac{8}{5}\right)=\frac{2}{5}\,.$$

Since, X and Y are identically distributed we have $E(X)=E(Y)$ and $Var(X)=Var(Y)$. From the marginal probability density function of X

$$E(X)=\int_{x=0}^{2} x\, f_X(x)\,dx=\frac{1}{4}\int_{x=0}^{2} x(x+1)\,dx=\frac{1}{4}\int_{x=0}^{2} (x^2+x)\,dx=\frac{1}{4}\left(\frac{x^3}{3}+\frac{x^2}{2}\right)_{x=0}^{2}=\frac{1}{4}\left(\frac{8}{3}+\frac{4}{2}\right)=\frac{7}{6}$$

and

$$E(X^2)=\int_{x=0}^{2} x^2\, f_X(x)\,dx=\frac{1}{4}\int_{x=0}^{2} x^2(x+1)\,dx=\frac{1}{4}\int_{x=0}^{2} (x^3+x^2)\,dx=\frac{1}{4}\left(\frac{x^4}{4}+\frac{x^3}{3}\right)_{x=0}^{2}=\frac{1}{4}\left(\frac{16}{4}+\frac{8}{3}\right)=\frac{5}{3}$$

and the variance becomes

$$Var(X)=E(X^2)-(E(X))^2=\frac{5}{3}-\frac{49}{36}=\frac{11}{36}.$$

In order to calculate the covariance, we need $E(XY)$ which is calculated as

$$E(XY)=\int_{y=0}^{2}\int_{x=0}^{2} xy\, f(x,y)\,dx\,dy=\frac{1}{8}\int_{y=0}^{2}\int_{x=0}^{2} xy(x+y)\,dx\,dy=\frac{1}{8}\int_{x=0}^{2}\left[\frac{x^2y^2}{2}+\frac{xy^3}{3}\right]_{y=0}^{2} dx$$

$$=\frac{1}{8}\int_{x=0}^{2}\left[2x^2+\frac{8x}{3}\right]dx=\frac{1}{8}\left[\frac{2x^3}{3}+\frac{8x^2}{6}\right]_{x=0}^{2}=\frac{4}{3}.$$

Thus the covariance between X and Y is calculated as

$$Cov(X,Y)=E(XY)-E(X)E(Y)=\frac{4}{3}-\left(\frac{7}{6}\right)\left(\frac{7}{6}\right)=\frac{4}{3}-\frac{49}{36}=\frac{48-49}{36}=-\frac{1}{36}$$

and the correlation between these two random variables is found to be

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}=\frac{-1/36}{\sqrt{(11/36)(11/36)}}=\frac{-1/36}{11/36}=-\frac{1}{36}.$$

This says that the random variables X and Y are negatively correlated. This means that when one increase the other decrease.

Finally let us calculate $E(Y \mid X=x)$. In order to calculate the conditional expectation we need to find the conditional probability density function. The conditinal probability density function of Y given $X=x$ is given by

$$f_{Y|X=x}(y\mid x)=\frac{f(x,y)}{f_X(x)}=\frac{(x+y)/8}{(x+1)/4}=\frac{x+y}{2(x+1)},\ 0<y<2.$$

That is,

$$f_{Y|X=x}(y\mid x)=\begin{cases}\dfrac{x+y}{2(x+1)} & , \quad 0<y<2\\ 0 & , \quad elsewhere\end{cases}$$

Note that

$$\int_{y=0}^{2} f_{Y|X=x}(y\mid x)dy=\frac{1}{2(x+1)}\int_{y=0}^{2}(x+y)dy=\frac{1}{2(x+1)}\left[xy+y^2/2\right]_{y=0}^{2}=1$$

and therefore,

$$E(Y \mid X = x) = \int_{y=0}^{2} y\, f_{Y|X=x}(y \mid x)dy = \frac{1}{2(x+1)} \int_{y=0}^{2} y(x+y)dy$$

$$= \frac{1}{2(x+1)}\Big[(x\,y^2/3) + y^3/3\Big]_{y=0}^{2} = \frac{2x+(8/3)}{2(x+1)} = \frac{3x+4}{3(x+1)}.$$

Example: Let X, Y and Z be three random variables with joint probability density function

$$f(x,y,z) = \begin{cases} c(x+y)z & , \quad 0 < x, y, z < 1 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

First, let us calculate the constant c. Since it is a joint probability density function the total probability has to be one. That is,

$$1 = \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} f(x,y,z)\,dz\,dy\,dx = c \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} (x+y)\,z\,dz\,dy\,dx = \frac{c}{2}$$

and thus $\boxed{c = 2}$. therefore, the joint probability density function of X, Y and Z is

$$f(x,y,z) = \begin{cases} 2(x+y)z & , \quad 0 < x, y, z < 1 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

Note also that the marginal probability density function of Z and the joint probability density function of (X,Y) can be obtained from

$$f_Z(z) = \int_{x=0}^{1} \int_{y=0}^{1} 2(x+y)\,z\,dy\,dx = 2z \;\; , \;\; f_{X,Y}(x,y) = \int_{z=0}^{1} 2(x+y)\,z\,dz = (x+y)$$

as

$$f_Z(z) = \begin{cases} 2z & , \quad 0 < z < 1 \\ 0 & , \quad \textit{elsewhere.} \end{cases} \quad \text{and} \quad f_{X,Y}(x,y) = \begin{cases} (x+y) & , \quad 0 < x, y < 1 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

Notice that since

$$f(x,y,z) = 2(x+y)z = (x+y)\; 2z = f_{X,Y}(x,y)\; f_Z(z)$$

the random variable Z is independent from (X,Y) and therefore $Cov(X,Z) = Cov(Y,Z) = 0$. The marginal probability density functions of X and Y are found by using the integrals

$$f_X(x) = \int_{y=0}^{1} f_{X,Y}(x,y)\,dy = \int_{y=0}^{1} (x+y)\,dy = x + \frac{1}{2} \quad \text{and} \quad f_Y(y) = \int_{y=0}^{1} (x+y)\,dx = y + \frac{1}{2}$$

as

$$f_X(x) = \begin{cases} x + \frac{1}{2} & , \quad 0 < x < 1 \\ 0 & , \quad \textit{elsewhere} \end{cases} \quad \text{and} \quad f_Y(y) = \begin{cases} y + \frac{1}{2} & , \quad 0 < y < 1 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

Moreover, the joint probability density functions of (X,Z) and (Y,Z) are

$$f_{X,Z}(x,z)=\begin{cases} z(2x+1) & , \quad 0<x,z<1 \\ 0 & , \quad elsewhere \end{cases} \quad \text{and} \quad f_{Y,Z}(y,z)=\begin{cases} z(2y+1) & , \quad 0<y,z<1 \\ 0 & , \quad elsewhere. \end{cases}$$

Since,

$$f_{X,Z}(x,z)=z(2x+1)=2z\,(x+(1/2))=f_Z(z)\,f_X(x)$$

The random variables X and Z are independent and thus $Cov(X,Z)=0$. Similarly,

$$f_{Y,Z}(y,z)=z(2y+1)=2z\,(y+(1/2))=f_Z(z)\,f_Y(y)$$

The random variables Y and Z are independent and thus $Cov(Y,Z)=0$.

In order to calculate the conditional expectations we need the conditional probability density functions. Let us try to calculate, $E(Z\mid X=x,Y=y)$ and $E(Y\mid X=x,Z=z)$. Since Z is independet from (X,Y) the conditional probability function of Z given $X=x$ and $Y=y$ is the same as $f_Z(z)$ because

$$f_{Z|X=x,Y=y}(z\mid x,y)=\frac{f(x,y,z)}{f_{X,Y}(x,y)}=\frac{2(x+y)\,z}{(x+y)}=2z$$

and thus,

$$E(Z\mid X=x,Y=y)=\int_{z=0}^{1} z\,f_{Z|X=x,Y=y}(z\mid x,y)\,dz=\int_{z=0}^{1} z\,f_Z(z)\,dz=\int_{z=0}^{1} 2z^2dz=\frac{2}{3}=E(Z).$$

On the other hand, in order to calculate $E(Y\mid X=x,Z=z)$ we need the conditional probability density function of Y given $X=x$ and $Z=x$ which can be obtained for $0<y<1$

$$f_{Y|X=x,Z=z}(z\mid x,y)=\frac{f(x,y,z)}{f_{X,Z}(x,z)}=\frac{2(x+y)\,z}{z\,(2x+1)}=\frac{2(x+y)}{(2x+1)}.$$

That is, the conditional probability density function is

$$f_{Y|X=x,Z=z}(z\mid x,y)=\begin{cases} \dfrac{2(x+y)}{(2x+1)} & , \quad 0<y<1 \\ 0 & , \quad elsewhere. \end{cases}$$

Note that this conditional probability density function does not depend on z and this is a probability density function because

$$\int_{y=0}^{1} f_{Y|X=x,Z=z}(z\mid x,y)\,dy=\int_{y=0}^{1}\frac{2(x+y)}{(2x+1)}dy=\frac{1}{(2x+1)}\int_{y=0}^{1} 2(x+y)dy=1.$$

Thus the conditional expectation

$$E(Y\mid X=x,Z=z)=\int_{y=0}^{1} y\frac{2(x+y)}{(2x+1)}dy=\frac{1}{(2x+1)}\int_{y=0}^{1} 2y\,(x+1)\,dy=\frac{3x+2}{2x+1}.$$

Remember the conditional probability of A given B can be written as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

and if the events A and B are independent we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) .$$

Now, using the properties of the conditional probability we calculate some probabilities as below:

$$P(Y < Z) = \int_{z=0}^{1} P(Y < Z \mid Z = z)\, f_Z(z)\, dz = \int_{z=0}^{1} P(Y < z)\, f_Z(z)\, dz = \int_{z=0}^{1} \left(\int_{y=0}^{z} f_Y(y) dy \right) f_Z(z)\, dz$$

$$= \int_{z=0}^{1} \left(\int_{y=0}^{z} [y + (1/2)]\, dy \right) 2z\, dz = \int_{z=0}^{1} \left(\frac{y^2 + y}{2} \right)_{y=0}^{z} 2z\, dz = \int_{z=0}^{1} (z^3 + z^2)\, dz = \frac{5}{12}$$

$$P(X > Y) = \int_{y=0}^{1} P(X > Y \mid Y = y) f_Y(y)\, dy = \int_{y=0}^{1} P(X > y)\, f_Y(y)\, dy = \int_{y=0}^{1} \left(\int_{x=y}^{1} f_X(x) dx \right) f_Y(y)\, dy$$

$$= \int_{y=0}^{1} \left(\int_{x=y}^{1} [x + (1/2)]\, dx \right) (y + 1/2)\, dy = \int_{y=0}^{1} \left(\frac{x^2 + x}{2} \right)_{x=y}^{1} (y + 1/2)\, dy$$

$$= \int_{y=0}^{1} \left(1 - \left(\frac{y^2 + y}{2} \right) \right) \left(y + \frac{1}{2} \right) dy = \frac{1}{2}.$$

Now, we want to calculate the following conditional probabilities:

$P(X > 1/2 | Z < 1/2)$ and $P(X > Y | Z < 1/2)$.

Note that since the random variables X and Z are independent

$$P(X > 1/2 | Z < 1/2) = \frac{P(X > 1/2\, ,\, Z < 1/2)}{P(Z < 1/2)} = \frac{P(X > 1/2)\, \cancel{P(Z < 1/2)}}{\cancel{P(Z < 1/2)}} = P(X > 1/2)$$

$$= \int_{x=1/2}^{1} f_X(x) dx = \int_{x=1/2}^{1} (x + 1/2) dx = \left(\frac{1}{2}(x^2 + x) \right)_{x=1/2}^{1} = \frac{5}{8}$$

and

$$P(X > Y | Z < 1/2) = \frac{P(X > Y)\ \cancel{P(Z < 1/2)}}{\cancel{P(Z < 1/2)}} = P(X > Y) = \frac{1}{2} .$$