However, without calculating the pdf of Z, $f_Z(z)$, it is possible to find μ_Z from the joint pmf or pdf of the X and Y random variables.

$$\mu_Z = E[Z]$$
$$= E[H(X,Y)]$$
$$= \begin{cases} \iint H(x,y)\, f_{XY}(x,y)\, dx\, dy & \text{, Continuous case} \\ \sum_j \sum_k H(x_j, y_k)\, p_{X,Y}(x_j, y_k) & \text{, Discrete case} \end{cases}$$

COVARIANCE, CORRELATION AND CORRELATION COEFFICIENT

(1) Covariance (σ_{XY}) between X and Y random variables is defined.

$$\sigma_{XY} = Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\, dx\, dy - \mu_X \mu_Y .$$

(2) CORRELATION

between X an Y random variable is defined as:

$$r_{XY} = E[XY]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\, dx\, dy.$$

(3) CORRELATION COEFFICIENT

between X and Y random variables is the normalized covarians and is defined as

$$\rho_{XY} = \frac{COV[X,Y]}{\sigma_X\ \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X\ \sigma_Y}$$

and $-1 \leq \rho_{XY} \leq 1.$

THEOREM:

(a) $COV[X,Y] = \sigma_{XY}$

$$= r_{XY} - \mu_X\, \mu_Y.$$

(b) $VAR[X+Y] = VAR[X] + VAR[Y] + 2\, COV[X,Y]$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\,\sigma_{XY}$$

(c) If $X = Y$,

$$\sigma_{XY} = \sigma_X^2 = \sigma_Y^2,$$

and

$$r_{XY} = E[X^2] = E[Y^2].$$

Definition:

Orthogonal Random Variables

Random variables X and Y are orthogonal if

$$r_{XY} = 0.$$

Definition:

Uncorrelated Random Variables

Random variables X and Y are uncorrelated if

$$\sigma_{XY} = 0.$$

Remark:

- This terminology is somewhat confusing, since orthogonal means zero correlation, while uncorrelated means zero covariance.
- On the other hand, we know that the correlation coefficient is closely related to the covariance of two random variables,

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \in [-1, 1].$$

- We correlation coefficient describes the information we gain about Y by observing X.
- For example, a positive correlation coefficient, $\rho_{XY} > 0$, suggests that when X high relative to its expected value, Y also tends to be high, and when X is low, Y is likely to be low.
- A negative correlation coefficient, $\rho_{XY} < 0$, suggests that a high value of X likely to be accompanied by a low value of Y and that a low value of X is likely to be accompanied by a high value of Y.

- A linear relationship between X and Y,

$$Y = aX + b$$

produces the extreme values

$$\rho_{XY} = \begin{cases} +1 & , \text{ for } a > 0 \\ -1 & , \text{ for } a < 0 \\ 0 & , \text{ for } a = 0 \end{cases}$$

2-D Jointly Gaussian Random Variables

- Let us assume

$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and

$Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, if they are independent random variables, the joint pdf can be written as follows;

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \forall x, \forall y.$$

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \frac{1}{\sqrt{2\pi}\,\sigma_Y} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}$$

$$-\infty < x, y < \infty$$

$$= \frac{1}{(2\pi)\,\sigma_X \sigma_Y} e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]}.$$

- However, if X and Y are correlated, we must introduce another parameter, ρ_{XY} correlation coefficient, or covariance between X and Y, σ_{XY}. Indeed, they are related:

$$\rho_{XY} \triangleq \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad \text{or}$$

$$\sigma_{XY} = \rho_{XY}\, \sigma_X\, \sigma_Y .$$

In this case,

$$(X,Y) \sim \mathcal{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY}).$$

There are five parameters instead of four.

- This correlation coefficient is introduced into the formula, $f_{XY}(x,y)$:

$$\frac{1}{2\pi \underbrace{\sigma_X \sigma_Y}} \longrightarrow \frac{1}{2\pi \underbrace{\sigma_X \sigma_Y \sqrt{1-\rho_{XY}^2}}}$$

$$-\frac{1}{2} \longrightarrow -\frac{1}{2} \frac{1}{\underbrace{(1-\rho_{XY}^2)}}$$

and substract from the parameters,

$$-2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)$$

and our jointly gaussian pdf will be

$$f_{XY}(x,y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1-\rho_{XY}^2}} e^{-\frac{1}{2(1-\rho_{XY}^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]},$$

for $-\infty < x, y < \infty$.

In order to understand the construction of this joint pdf, let us consider n-dimensional case. The joint pdf of $X_1, X_2, \dots, X_n$ random variables,

$\underline{\mu}_{\underline{X}}$ shows the average value vector

$$\underline{\mu}_{\underline{X}} = \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$ and

the $\underline{C}_{\underline{X}\underline{X}}$ covariance matrix that is defined by

( )^T denotes the transpose of a vector

$$\underline{C}_{\underline{X}\underline{X}} = E\left[(\underline{X} - \underline{\mu}_{\underline{X}})(\underline{X} - \underline{\mu}_{\underline{X}})^T \right]$$

$$= E\left[\begin{bmatrix} X_1 - \mu_{X_1} \\ \vdots \\ X_n - \mu_{X_n} \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 & \dots & X_n - \mu_n \end{bmatrix} \right]$$

$$= \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} & \dots & \sigma_{X_1X_n} \\ \sigma_{X_2X_1} & \sigma_{X_2}^2 & \dots & \sigma_{X_2X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_nX_1} & \sigma_{X_nX_2} & \dots & \sigma_{X_n}^2 \end{bmatrix}.$$

The covariance matrix $\underline{C}_{\underline{X}\underline{X}}$ is a symmetric matrix since $Cov(X_i, X_j) = Cov(X_j, X_i)$.

Then,

$$\underline{X} = (X_1, X_2, \dots, X_n) \sim \mathcal{N}(\underline{\mu}_{\underline{X}}, \underline{C}_{\underline{X}\underline{X}})$$

and joint pdf is given by

$$f_{\underline{X}}(\underline{x}) = f_{X_1, X_2, \dots, X_2}(x_1, x_2, \dots, x_n)$$

$$= \frac{1}{(2\pi)^{n/2} |\underline{C}_{\underline{X}\underline{X}}|^{1/2}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_{\underline{X}})^T \underline{C}_{\underline{X}\underline{X}}^{-1} (\underline{x}-\underline{\mu}_{\underline{X}})}$$

$$-\infty \leq x_i \leq \infty \quad , i = 1, \dots, n.$$

where

$|\underline{C}_{\underline{X}\underline{X}}|$ is the determinant of the Covariance matrix.

- The covariance matrix for the 2-D case is given by

$$\underline{C}_{\underline{X}\underline{X}} = E\left[\underbrace{\begin{pmatrix} X_1 - \mu_{X_1} \\ X_2 - \mu_{X_2} \end{pmatrix}}_{(\underline{X}-\underline{\mu}_{\underline{X}})} \underbrace{\begin{pmatrix} X_1 - \mu_{X_1} & X_2 - \mu_{X_2} \end{pmatrix}}_{(\underline{X}-\underline{\mu}_{\underline{X}})^T} \right]$$

$$\underline{C}_{\underline{X}\underline{X}} = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{X_1}^2 & \rho_{XY}\,\sigma_{X_1}\sigma_{X_2} \\ \rho_{XY}\,\sigma_{X_1}\sigma_{X_2} & \sigma_{X_2}^2 \end{bmatrix}$$

and

$$\underline{C}_{\underline{X}\underline{X}}^{-1} = \frac{1}{\sigma_{X_1}^2\,\sigma_{X_2}^2\,(1-\rho_{XY}^2)} \begin{bmatrix} \sigma_{X_2}^2 & -\rho_{XY}\sigma_{X_1}\sigma_{X_2} \\ -\rho_{XY}\sigma_{X_1}\sigma_{X_2} & \sigma_{X_1}^2 \end{bmatrix}.$$

Remark:

If $X_1, X_2, \ldots, X_n$ are independent Gaussian random variables, the covariance matrix is diagonal

$$\underline{C}_{\underline{X}\underline{X}} = \begin{bmatrix} \sigma_{X_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{X_2}^2 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & \sigma_{X_n}^2 \end{bmatrix}$$

• all covariances are zero.

Therefore,

$$f_{\underline{X}}(\underline{x}) = f_{X_1}(x_1)\, f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}}\, e^{-\frac{1}{2\sigma_i^2}(x_i - \mu_{X_i})^2}$$

all $x_1, x_2, \ldots, x_n$.

Marginal Probability Distributions

If $(X,Y) \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$

2-D normal distribution with a probability density function, $f_{XY}(x,y)$, then the marginal probability distributions of X and Y are normal

$$X \sim N(\mu_X, \sigma_X^2)$$

$$Y \sim N(\mu_Y, \sigma_Y^2)$$

and

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

$$= \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}, \quad -\infty < x < \infty$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

$$= \frac{1}{\sigma_Y \sqrt{2\pi}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}, \quad -\infty < y < \infty.$$

On the other hand, from the joint probability density function, the covariance of X and Y random variables can be calculated:

$$\sigma_{XY} = \int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)\,dx\,dy$$

and the correlation coefficient is written as,

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Conditional Distributions of 2-D (bivariate) Normal Random Variables

- The conditional probability distribution of Y given $X=x$ is given by

$$f_{Y|x}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$= \frac{1}{\sigma_Y \sqrt{2\pi} \sqrt{1-\rho_{XY}^2}} e^{-\frac{1}{\underbrace{2\sigma_Y^2(1-\rho_{XY}^2)}_{\sigma_{Y|x}^2}} \left[y - \underbrace{\left[\mu_Y + \rho_{XY} \frac{\sigma_Y}{\sigma_X}(x-\mu_X) \right]}_{\mu_{Y|x}} \right]^2}$$

$$-\infty < y < \infty$$

and similarly the conditional pdf of X given $Y=y$ can be written [see the texbook].

- From these conditional pdfs, we can make some observations or remarks:

(1) Conditional pdfs are normal distributed:

$$f_Y(y|x) \sim \mathcal{N}(\mu_{Y|x}, \sigma_{Y|x}^2)$$

$$f_X(x|y) \sim \mathcal{N}(\mu_{X|y}, \sigma_{X|y}^2).$$

- From this expression, we can obtain conditional mean and variance:

$$\mu_{Y|x} = \mu_Y + \rho_{XY} \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$$

$$\mu_{X|y} = \mu_X + \rho_{XY} \frac{\sigma_X}{\sigma_Y} (y - \mu_Y)$$

and

$$\sigma_{Y|x}^2 = \sigma_Y^2 (1 - \rho_{XY}^2)$$

$$\sigma_{X|y}^2 = \sigma_X^2 (1 - \rho_{XY}^2).$$

From these results, we can see that conditional variances are constants. However, conditional mean values depend upon the given values $X = x$ or $Y = y$.

(2) In order to determine the effect of The correlation coefficient, we set $\rho_{XY} = 0$, we get the conditional pdfs as follows;

In other words, for $\rho_{XY}=0$,

$$f_{Y|x}(y|x) = f_Y(y) \text{ , for all } x,y$$

and

$$f_{X|y}(x|y) = f_X(x) \text{ , for all } x,y.$$

At this situation, X and Y random variables are independent, namely,

$$f_{XY}(x,y) = f_X(x) f_Y(y) .$$

- For jointly normal distributed (X,Y)

$$\rho_{XY}=0 \;\underset{\text{always}}{\overset{\text{implies}}{\rightleftarrows}}\; X \text{ and } Y \text{ independent}$$

"For only Gaussian distribution".

Other distributions for $\rho_{XY}=0$, it is only necessary to be independent.

(3) If $|\rho_{XY}| \longrightarrow 1$, the conditional variances approaches to zero, namely,

$$|\rho_{XY}| \to 1 \Rightarrow \sigma_{Y|x}^2 \to 0 \text{ and } \sigma_{X|y}^2 \to 0.$$

At this case, conditional pdfs tend to be impulse functions,

$$f_Y(y|x) = \delta(y - \mu_{Y|x})$$

given the value $X=x$, the value of random variable Y 100 % is determined.

Similar result is obtained for random variable X,

$$f_X(x|y) = \delta(x - \mu_{X|y}).$$

i.e. the relationship is linear, $Y=aX+b$.

Example: The midterm exam results of a course given at a university, are found to be jointly normal distributed:

X = the first exam results

Y = the second exam results.

The distribution parameters are given by

$$(X,Y) \sim \mathcal{N}\left(\mu_X = 75,\ \mu_Y = 83,\ \sigma_X^2 = 25,\ \sigma_Y^2 = 16,\ \rho_{XY} = 0.8\right).$$

Question:

If the grade of the first exam is 80 of a student, what is the probability of getting greater than 80 for the second exam?

First, we must calculate the conditional mean and the conditional variance of this conditional pdf:

$$f_{Y|x=80}(y|80) = \frac{1}{\sigma_{Y|80}\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma_{Y|80}^2}(x-\mu_{Y|80})^2}$$

Conditional standard deviation

Conditional Variance

Conditional Mean.

$$\mu_{Y|x=80} = \mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(80-\mu_X)$$

$$= 83 + 0.8\,\frac{4}{5}(80-75)$$

$$= 86.2$$

and

$$\sigma^2_{Y|x=80} = \sigma_Y^2\left(1-\rho_{XY}^2\right)$$

$$= 16\left(1-0.8^2\right)$$

$$= 5.76 .$$

Therefore,

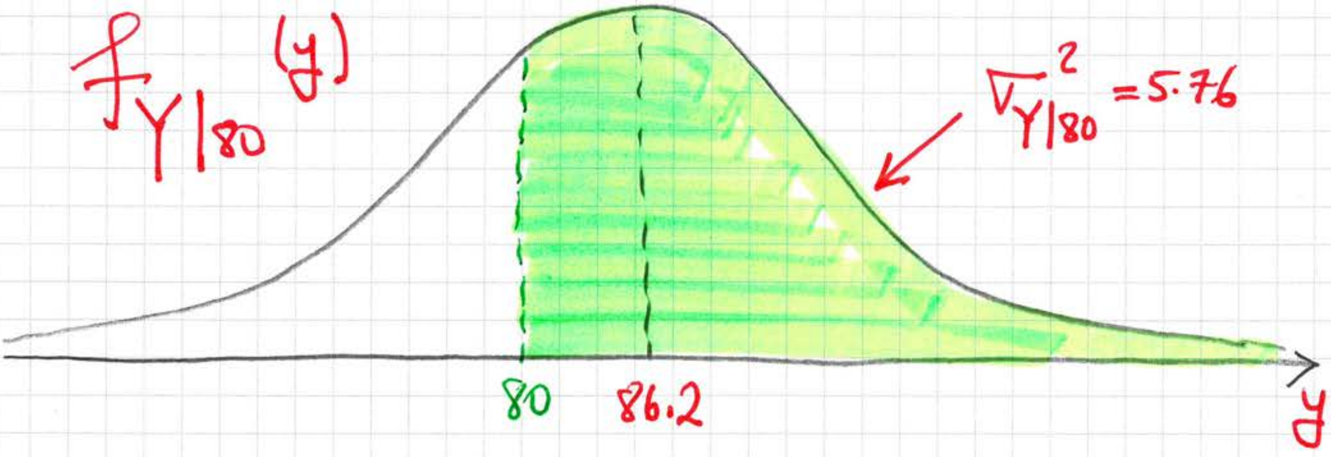
$$Y \mid x=80 \longrightarrow \mathcal{N}(86.2, 5.76)$$

and

$$f_{Y|80}(y) = \frac{1}{\sqrt{2\pi(5.76)}}\, e^{-\frac{1}{2(5.76)}(x-8.76)^2}$$

Subject Name Probability and Statistics

(4)



$$Pr\{Y > 80 \mid x=80\} = 1 - Pr\{Y \leq 80 \mid x=80\}$$

$$= 1 - \Phi\left(\frac{80-86.2}{\sqrt{5.76}}\right)$$

$$= 1 - \Phi(-2.58)$$

$$= 1 - [1 - \Phi(2.58)]$$

$$= \Phi(2.58) = 0.9951.$$

* $\Phi(z)$ is the standard normal distribution and the value $z=2.58$ is found at the table.

Remark:

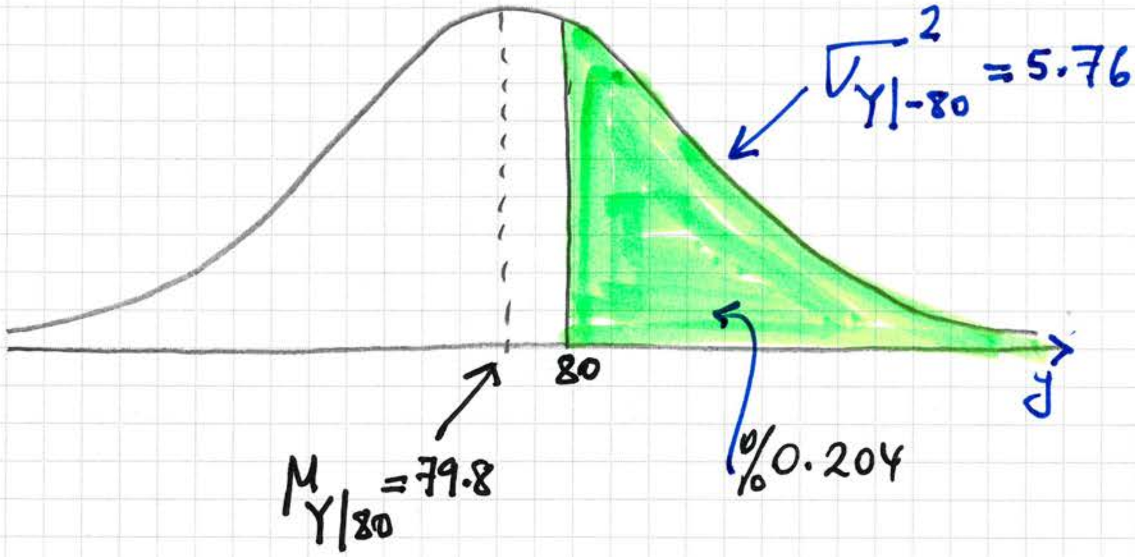
If we change the sign of the correlation coefficient, for $\rho_{XY} = -0.8$, we must calculate the new conditional average value,

$$\mu_{Y|80} = 80 - 0.8\,\frac{4}{5}(80-75) = 79.8$$

$$\sigma_{Y|80} = \sigma_Y^2(1-\rho_{XY}^2) = 5.76.$$

Since we take the square value of the correlation, the conditional variance does not change.

$$f_{Y|-80}(y) = \frac{1}{\sqrt{2\pi(5.76)}} e^{-\frac{1}{2(5.76)}(y-79.8)^2}$$



$$\Pr\{Y > 80 \mid x = 80\} = 1 - \Phi\left(\frac{80 - 79.8}{\sqrt{5.76}}\right)$$

$$= 1 - \Phi(0.083)$$

$$= 1 - 0.796$$

$$= 0.204.$$

Another words, if you are getting a good grade at the first exam, the probability if getting good grade is reducing for $\rho_{xy} < 0$.