Topic 10: Covariance and More on Conditional Expectation

Outline

- Covariance (共變異數) and Correlation (相關性)
- Conditional Expectation
- Total Expectation Theorem/ Total Variance
- Bivariate Normal or Jointly Gaussian (雙變量常態分佈)

Reading: Chap. 4.2 ~ Chap. 4.3

Covariance and Correlation

Definition:

The covariance of two random variables X and Y is denoted by cov(X, Y), and is defined by

$$\text{cov}(X,Y) \triangleq E\left[(X - E[X])(Y - E[Y])\right]$$

When cov(X, Y) = 0, we say that X and Y are uncorrelated.

$$\text{cov}(X,Y) = 0 \quad \Leftrightarrow \quad E[XY] = E[X]E[Y]$$

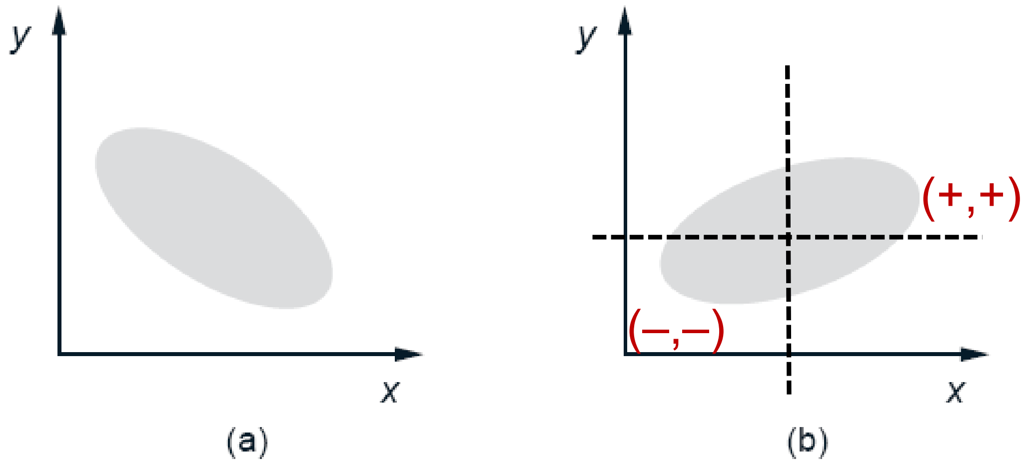
Uncorrelated 中文翻成“零相關”為妥 (而非“不相關”) ，表示covariance=0

Question

Uncorrelated RVs 的物理意義為何? 和 independent 的差異何在?

Interpretation of Covariance

- Covariance provides a qualitative indicator of the relation between two random variables
  - When cov(X,Y)>0, X-$E[X]$ and Y-$E[Y]$ "tend" to have the same sign
  - When cov(X,Y)<0, X-$E[X]$ and Y-$E[Y]$ "tend" to have the opposite sign



兩隨機變數正相關表示(平移至期望值後)同號的傾向較大，如上圖(b)

正相關表示：若 X-$E[X]$ 為正，則 Y-$E[Y]$ 也為正的機會較高

Uncorrelated and Independent

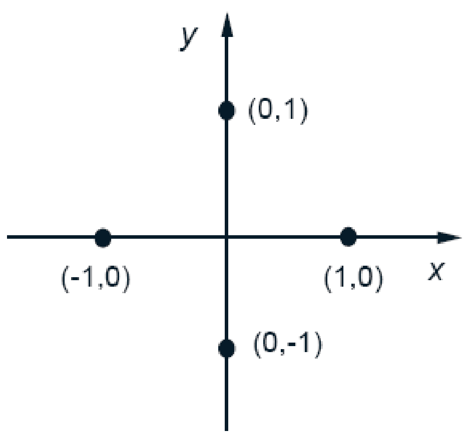
重要:

Independence implies uncorrelatedness. But the converse is generally **NOT** true.

$$
\begin{aligned}
&\text{Independent} \quad \Rightarrow \quad f_{XY}(x,y) = f_X(x) f_Y(y) \\
&\text{uncorrelated} \quad \Rightarrow \quad E[XY] = E[X]E[Y]
\end{aligned}
$$

上述兩式並非等義!

Example: (Uncorrelatedness does NOT imply independence)



Ex.4.13 (p.218) Each of the four points shown has probability 1/4. Here X and Y are uncorrelated but not independent.

Correlation Coefficient

The correlation coefficient ρ of two random variables X and Y that have nonzero variances is defined as

$$\rho(X, Y) \triangleq \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}$$

- A normalized version of the covariance
- We must have $-1 \leq \rho \leq 1$ (p.250: problem 20 and 21)
- If ρ >0, $X - E[X]$ and $Y - E[Y]$ "tend" to have the same *sign* , *i.e.* $(+, +)$ or $(-, -)$
- The value of $|\rho|$ provides a normalized measure of the extent to the *relation* between $X - E[X]$ and $Y - E[Y]$

Example:

Let $Y - E[Y] = c\ (X - E[X])$ for a positive constant c. Then, the correlation coefficient $\rho = 1$. This says that $X - E[X]$ exactly aligns with $Y - E[Y]$, subject to a positive constant (STRONG correlation)

Variance of the Sum of Random Variables

Several properties related to covariance:

$$\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$$

$$\mathrm{cov}(X,X) = \mathrm{var}\,(X)$$

$$\mathrm{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}(X_i, Y_j)$$

Variance of the Sum of Random Variables

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{var}(X_i) + 2\sum\sum_{i<j} \mathrm{cov}(X_i, X_j)$$

- For $n=2$: var(X_1+X_2)=var(X_1) +2cov(X_1,X_2)+var(X_2)
- See textbook page 221 for an example.

Example

Consider the problem that n people throw their hat in a box and then pick a hat at random. Let the number of people who pick their own hat be denoted by X. Find var(X). First, define the indicator X_i as follows:

$$X_i = \begin{cases} 1, & \text{if } i\text{th person picks the correct hat} \\ 0, & \text{otherwise.} \end{cases}$$

We know P(X_i=1)=1/n , so

$$E[X_i] = \frac{1}{n}, \quad \mathrm{var}(X_i) = \frac{1}{n}\left(1 - \frac{1}{n}\right)$$

$$\mathrm{cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$$

Then we have $X = X_1 + X_2 + \ldots + X_n$

$$\mathrm{var}(X) = \sum_{i=1}^{n} \mathrm{var}(X_i) + \sum_{i \neq j} \mathrm{cov}(X_i, X_j)$$

Conditional Expectation and Total Expectation

Recall: The conditional expectation $E[X \mid Y=y]$ is defined by

$$
\begin{aligned}
E[X|Y=y] &= \sum_x x p_{X|Y}(x|y), \quad \text{discrete case,} \\
E[X|Y=y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx, \quad \text{continuous case.}
\end{aligned}
$$

It should be noted that $E[X \mid Y]$ depends on Y, and is therefore a random variable whose value is $E[X \mid Y=y]$ when the outcome of Y is y.

Law of Iterated Expectations: (Total Expectation Theorem, p. 104, p. 173)

$$
E\big[E[X \mid Y]\big] = \left\{ \begin{array}{c} \sum_y E[X \mid Y=y] p_Y(y) \\ \int_{-\infty}^{\infty} E[X \mid Y=y] f_Y(y)\,dy \end{array} \right\} = E[X]
$$

$$
\boxed{E[E[X|Y]] = E[X]}
$$

Example of $E[\ E[X \mid Y]\ ]=E[X]$

Orthogonality Principle:

Show that $E\Big[(X - E[X|Y]) \cdot g(Y)\Big] = 0$ for any function $g(.)$.

(Proof)

We first need to show $E\big[Xg(Y)|Y\big] = g(Y)E[X|Y]$ (See problem 25.)

Conditional Variance

The conditional distribution of X given $Y = y$ has a mean $E[X \mid Y = y]$, and by the same token, it also has a variance defined as

$$\text{var}(X|Y = y) \triangleq E\left[(X - E[X|Y = y])^2 \mid Y = y\right]$$

1. Var($X|Y$) is a function of Y, and is therefore a random variable
2. Law of total variance

$$\text{var}(X) = E[\text{var}(X|Y)] + \text{var}(E[X|Y])$$

Proof of Law of Total Variance

$$\text{var}(X) = E[\text{var}(X|Y)] + \text{var}(E[X|Y])$$

(a) Recall: $\text{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$

(b) $\text{var}(X \mid Y) = \mathbf{E}[X^2 \mid Y] - (\mathbf{E}[X \mid Y])^2$

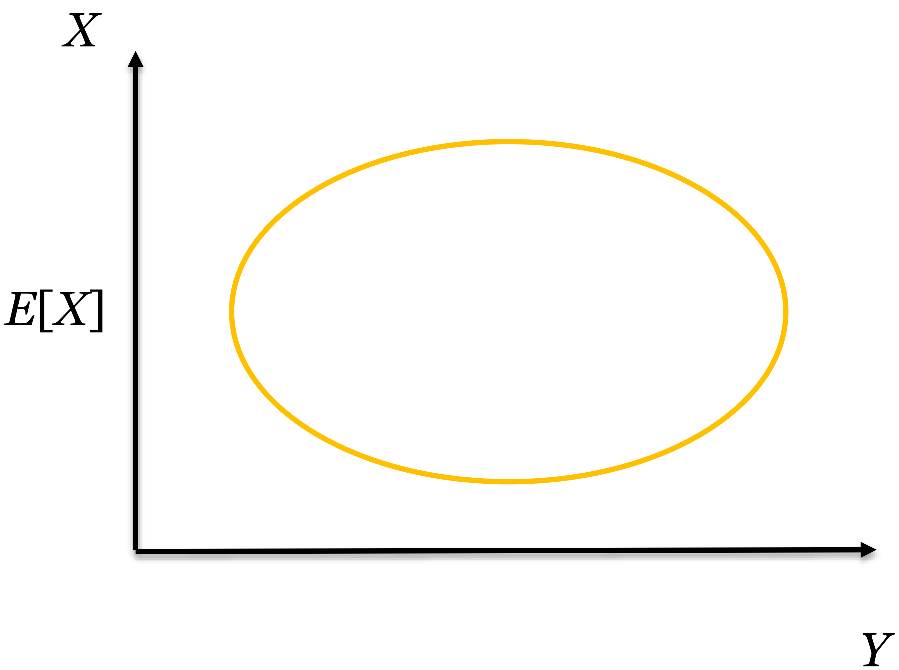
(c) $\mathbf{E}[\text{var}(X \mid Y)] = \mathbf{E}[X^2] - \mathbf{E}[\,(\mathbf{E}[X \mid Y])^2\,]$

(d) $\text{var}(\mathbf{E}[X \mid Y]) = \mathbf{E}[\,(\mathbf{E}[X \mid Y])^2\,] - (\mathbf{E}[X])^2$

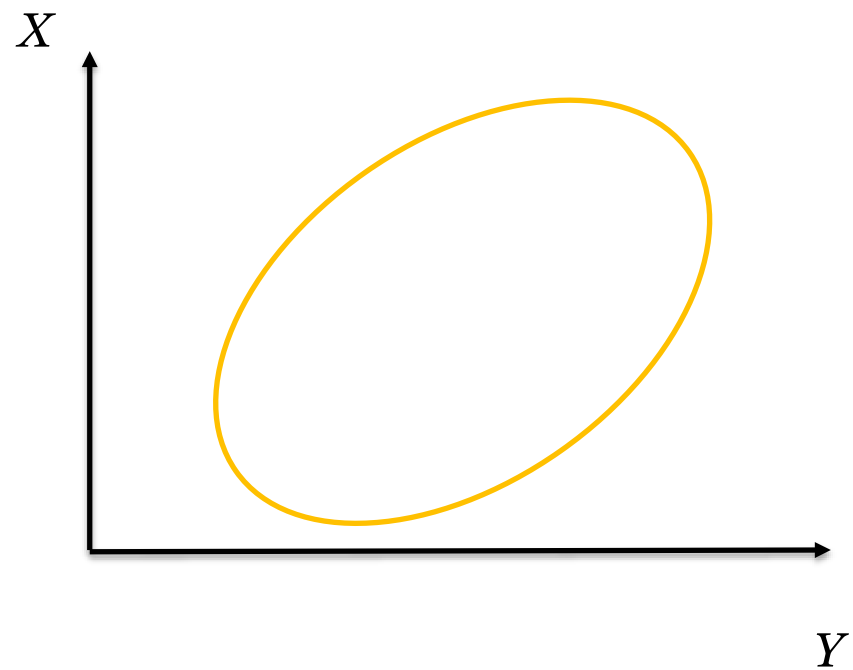
Sum of right-hand sides of (c), (d):
$\mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \text{var}(X)$

Intuitions behind Law of Total Variance

$$\mathrm{var}(X) = E[\mathrm{var}(X|Y)] + \mathrm{var}\left(E[X|Y]\right)$$



When $\mathrm{E}[X|Y] = \mathrm{E}[X]$, a constant

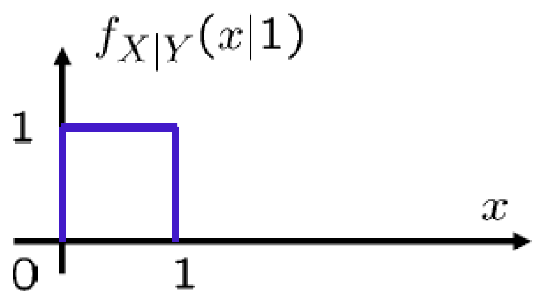


When $\mathrm{E}[X|Y]$ depends on Y

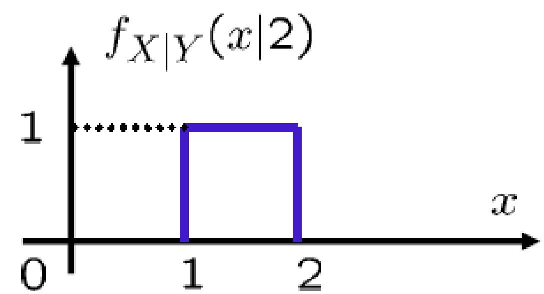
Example (1)

- Throw a biased coin, with $\mathrm{P}(\mathrm{H}) = \frac{1}{3}$:
  - If H: choose a number uniformly in [0, 1].
  - If T: choose a number uniformly in [1, 2].

- Using random variables:



$p_Y(1) = \mathbf{P}(Y = 1) = \frac{1}{3}$

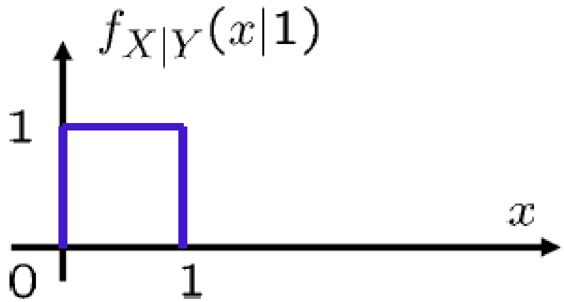


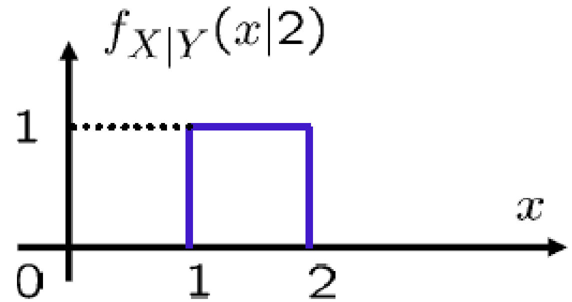
$p_Y(2) = \mathbf{P}(Y = 2) = \frac{2}{3}$

- We're interested in $\mathrm{Var}(X)$.

Example (2)

$$p_Y(1) = \frac{1}{3}$$
$$p_Y(2) = \frac{2}{3}$$





- Use: $\mathrm{Var}(X) = \mathbf{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbf{E}[X|Y])$

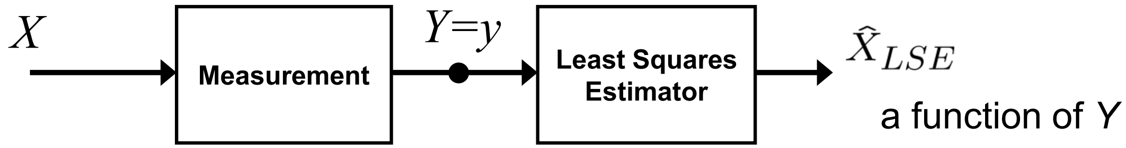
$$\left.\begin{array}{l} \mathrm{Var}(X|Y=1) = 1/12 \\ \mathrm{Var}(X|Y=2) = 1/12 \end{array}\right\} \Rightarrow \mathbf{E}[\mathrm{Var}(X|Y)] = 1/12$$

$$\left.\begin{array}{l} \mathbf{E}[X|Y=1] = 1/2 \\ \mathbf{E}[X|Y=2] = 3/2 \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[X] = 7/6 \\ \mathbf{E}[\mathbf{E}[X|Y]^2] = 19/12 \end{array}\right.$$

$$\mathrm{Var}(\mathbf{E}[X|Y]) = \mathbf{E}[\mathbf{E}[X|Y]^2] - \mathbf{E}[\mathbf{E}[X|Y]]^2 = 2/9$$

- Finally: $\mathrm{Var}(X) = 1/12 + 2/9 = 11/36$

Least Squares Estimation/Minimum Mean Square Error (MMSE)



Observed (Measured, Received)
Typically continuous

In many practical applications, we want to form an estimate of the value of a random variable X given the value of a related random variable Y, which may be viewed as some form of "measurement" of X. (For example, estimate the distance of an object from a radar)

A popular formulation is based on finding the estimate c, *based on the observed Y*, that *minimizes* the expected value of the squared error $(X\text{-}c)^2$.

$$\hat{X}_{LSE} = \arg\min_{c} E[(X-c)^2]$$

This is called the *least squares estimation* in our textbook, or
minimum mean-square error (MMSE) estimation in a large body of literature

MMSE Estimator

The problem here is to find a function $g(Y)$ such that the mean-square error (MSE) is minimized:

$$\hat{X}_{LSE} = \arg\min_{g(Y)} E\left[\left(X - g(Y)\right)^2\right]$$

It can be shown (Section 8.3) that out of all possible estimators $g(Y)$, the MSE is minimized when

$$g(Y) = E[X|Y]$$

That is

$$E\left[\left(X - E[X|Y]\right)^2\right] \le E\left[\left(X - g(Y)\right)^2\right] \quad \text{for all functions } g(Y)$$

➢ This can be proved using the othogonality principle (page 9).

Example

1. An IoT sensor is probing the temperature X on the top of Mt. Jade. It is known that the measurement of the IoT sensor is corrupted by Gaussian noise Z, where the measured signal Y is modeled by
$$Y = X + Z,$$
where $X \sim N(0, \sigma_X^2)$ and $Z \sim N(0, \sigma_Z^2)$ are independent normal RVs. What is the MMSE estimate of X provided the sensor has observed Y?

Jointly Normal Random Variables (p.254; problem 28*)

Two random variables X and Y are said to be jointly normal if they can be expressed in the form

$$\begin{aligned} X &= aU + bV, \\ Y &= cU + dV. \end{aligned}$$

where a, b,c and d are some scalars, U and V are independent normal random variables.

Remarks:

1. (Recall) Any linear combination of U and V is Gaussian. So, X or Y is Gaussian
2. If X and Y are jointly Gaussian, then ***any*** linear combination $Z=s_1X+s_2Y$ of X and Y has a Gaussian distribution (這是另外一種定義bivariate normal的方式)
3. If X and Y are jointly normal, then either X or Y is marginally (individually) normal. However, the *converse is in general not true*.

Uncorrelated Jointly Normal RVs Are Independent

A VERY important property for jointly normal random variables:

If two random variables X and Y are jointly normal and are uncorrelated, then they are independent. (See page 20 of this topic.)

Note: Generally, two uncorrelated random variables are not necessarily independent. (See page 4 of this topic.)

Example

If X and Y are zero mean jointly normal with variances σ_X^2 and σ_Y^2 respectively, find a constant α such that $X - \alpha Y$ and Y are independent.

(Sol.)

Conditional Distribution of Jointly Normal RVs

Question

Let X and Y have a bivariate (jointly) normal distribution (Assume zero mean, for simplicity). Then, what is the conditional density $f_{X|Y}(x|y)$ of X given Y

Two approaches: 1) Use the conditional density formula; or 2) Decompose X into $X=X-\alpha Y+\alpha Y$, such that $X-\alpha Y$ and Y are indep. (X and Y are marginal normal)

We adopt the 2nd approach here:

1. From last page's example, the conditional expectation is given by

$$\boxed{E[X|Y] = E[X] + \rho\frac{\sigma_X}{\sigma_Y}\Big(Y - E[Y]\Big)} \qquad (1)$$

   It is a linear function of Y and therefore has a normal PDF
2. The error $X\text{-}E[X|Y]$ is zero mean, normal, and independent of Y, with variance
   $(1\text{-}\rho^2)\sigma_X^2$
3. The conditional distribution of X given Y is normal with mean $E[X|Y]$ (given by the above eqn. (1)) and variance $(1\text{-}\rho^2)\sigma_X^2$

Jointly Normal PDF

Assume zero mean again for simplicity. From the previous page, the conditional distribution of X given Y is normal with mean $E[X|Y]$ and variance $(1-\rho^2)\sigma_X^2$

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_X} \cdot e^{-\frac{\left(x-\rho\frac{\sigma_X}{\sigma_Y}y\right)^2}{2(1-\rho^2)\sigma_X^2}}$$

We can find the joint PDF $f_{X,Y}(x,y)$ of X and Y using the multiplication rule

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

which takes the form $f_{X,Y}(x,y) = c \cdot e^{-q(x,y)}$

where $$c = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_X\sigma_Y} \qquad q(x,y) = \frac{\frac{x^2}{\sigma_X^2} - 2\rho\frac{xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2}}{2(1-\rho^2)}$$

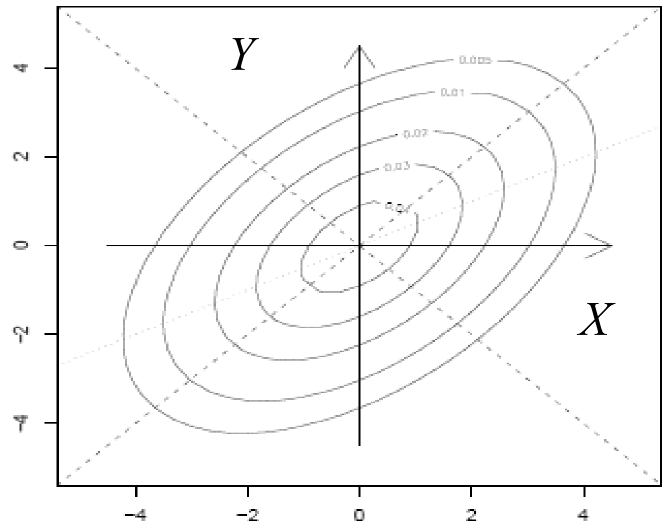
Note:

When $\rho=0$, X and Y are uncorrelated. And the above PDF also reveals that X and Y are independent with $\rho=0$ (since joint PDF factors)

Jointly Normal PDF Contours

$$f_{X,Y}(x,y) = c \cdot e^{-q(x,y)}$$

$$q(x,y) = \frac{\frac{x^2}{\sigma_X^2} - 2\rho\frac{xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2}}{2(1-\rho^2)}$$



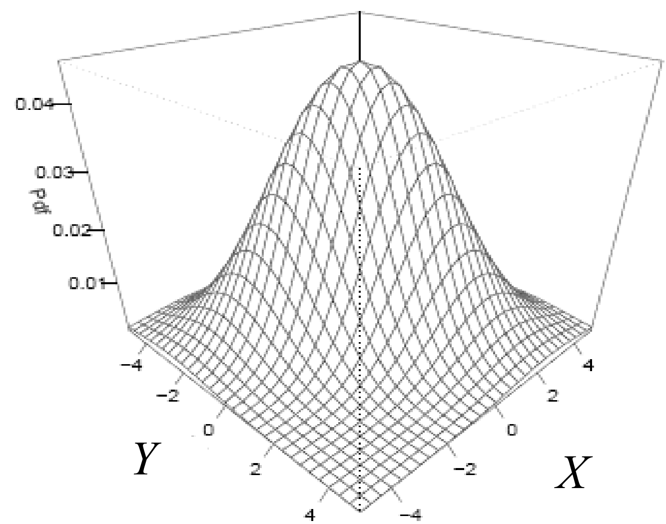
ρ=0.5

- Contour (等高線、橫切面圖)

$f_{X,Y}(x,y) = \text{constant } C_1$

➔ $q(x,y) = \text{constant } C_2$

- What about a vertical slice section at $Y=y$? (縱切面圖)



Example

1. An IoT sensor is probing the temperature X on the top of Mt. Jade. It is known that the measurement of the IoT sensor is corrupted by Gaussian noise Z, where the measured signal Y is modeled by

$$Y = X + Z,$$

where $X \sim N(0, \sigma_X^2)$ and $Z \sim N(0, \sigma_Z^2)$ are independent normal RVs. What is the MMSE estimate of X provided the sensor has observed Y?