743- Regression and Time Series

Mamikon S. Ginovyan

© 2012 Mamikon Ginovyan

The Prediction Problem

© 2012 Mamikon Ginovyan

1. Probabilistic Approach

- **Examples**

1. A stock-holder wants **to predict** the value of his holdings at some time in the **future** on the basis of his **past** experience with the market and his portfolio.
2. A meteorologist wants to estimate the amount of rainfall in the coming spring.
3. A government expert wants to predict the amount of heating oil needed next winter.

- **The frame we shall fit these and similar problems into is the following problem, called Prediction Problem.**

© 2012 Mamikon Ginovyan

The Prediction Problem (PP)

- Suppose we have some information represented by a RV ***X***, or by a random vector $\underline{X} = (X_1, ..., X_n)$, and we want to **predict (estimate)** the value of some quantity represented by a RV ***Y***, using the **information contained** in ***X***, that is, we want to find a function $g(\cdot)$ defined on the range of ***X*** (or $\underline{X}$) such that the RV

$$\hat{Y} = g(X), \quad or \quad \hat{Y} = g(\underline{X})$$

is **"close"** to ***Y***, then

$\hat{Y}$ is called the **predictor (or estimator)** for ***Y*;**

$Y - \hat{Y}$ is called the **prediction error**.

© 2012 Mamikon Ginovyan

The Prediction Problem (PP)

- **It is clear that**

a) we need to have some information about the **joint distribution** of X and Y, and

b) we must specify the **"measure of closeness"**.

- There are **different measures** of the "**closeness**" of $\hat{Y}$ to Y (distances between $\hat{Y}$ and Y), and

the **best predictor will depend on the measure** chosen.

© 2012 Mamikon Ginovyan

Measure of Closeness

- Two common used measures are

(a)
$$\left(\hat{Y}-Y\right)^2 = \left(g(X)-Y\right)^2$$
= the **squared error**

= the **quadratic loss** function;

(b)
$$\left|\hat{Y}-Y\right| = \left|g(X)-Y\right|$$
= the **absolute error**

= the **absolute loss** function.

© 2012 Mamikon Ginovyan

Measure of Closeness

- Since ***X*** and ***Y*** are RV's the distances $\left(g(X)-Y\right)^2$ and $\left|g(X)-Y\right|$ as functions of RV's will also be **RV's**.

So we need to **take expectations** and as **measures of closeness** of $\hat{Y}$ to ***Y*** consider the functions:

(a') $E[\left(\hat{Y}-Y\right)^2]=E[\left(g(X)-Y\right)^2]$

= the **mean squared error (MSE)**

= the **quadratic risk** function.

(b') $E\left|\hat{Y}-Y\right|=E\left|g(X)-Y\right|$

= the **mean absolute error**

= the **absolute value risk** function.

© 2012 Mamikon Ginovyan

Best Predictor (Special Case)

We begin the search for the **best predictor** $\hat{Y} = g(X)$ in the sense of minimizing

$$MSE = E\left(\hat{Y} - Y\right)^2$$

by considering the **special-trivial** case in which $\boldsymbol{X}$ is a **constant** (**non-random**), that is, $\boldsymbol{X} = \boldsymbol{x}$.

(**This case is important for Regression Theory**).

© 2012 Mamikon Ginovyan

Best Predictor (Special Case)

- In this **special case** all the predictors

$$\hat{Y} = g(X) = g(x) = c$$

are **constant** and the best one is that number $c_0 = g(x_0)$, which **minimizes** the **MSE**:

$$MSE = E\left(Y - c\right)^2$$

as a function of $\boldsymbol{c}$, that is,

$$E\left(Y - c_0\right)^2 = \min_c E\left(Y - c\right)^2 .$$

© 2012 Mamikon Ginovyan

Best Predictor (Special Case)

Theorem 1.

Let $R(c) = E(Y-c)^2$.

Then either

(a) $R(c) = \infty$ for all ***c***, or

(b) $R(c) < \infty$ and ***R(c)*** is **minimized uniquely** by $c_0 = E(Y)$, that is,

$$E(Y-EY)^2 = \min_c E(Y-c)^2 .$$

So, the **best predictor** in this case is the **mean** of ***Y***: $\hat{Y} = E(Y)$.

© 2012 Mamikon Ginovyan

Theorem 1-Proof

- **Proof.** Whatever ***Y*** and ***c***, we have

$$\frac{1}{2}Y^2 - c^2 \le (Y-c)^2 = Y^2 - 2cY + c^2 \le 2(Y^2 + c^2).$$

Hence (taking expectation)

$$\frac{1}{2}R(0) - c^2 \le R(c) \le 2\left[R(0) + c^2\right].$$

Therefore

$R(c) = \infty$ for all ***c*** **unless** $R(0) < \infty.$

If $R(0) < \infty,$ then $E(Y^2) < \infty$ and we can write

$$R(c) = E(Y^2) - 2cE(Y) + c^2.$$

© 2012 Mamikon Ginovyan

Solutions of the minimum problem

- **Probabilistic solution of the minimum problem.**

$$R(c) = E\left(Y - c\right)^2 = E(Y^2) - 2cE(Y) + c^2$$
$$= \left\{E(Y^2) - [E(Y)]^2\right\} + \left\{[E(Y)]^2 - 2cE(Y) + c^2\right\}$$
$$= Var(Y) + [E(Y) - c]^2.$$

Since both terms on the right are **non-negative**, we see that $R(c)$

has a **unique minimum** (equal to $Var(Y)$) at $c_0 = E(Y)$.

© 2012 Mamikon Ginovyan

Solutions of the minimum problem

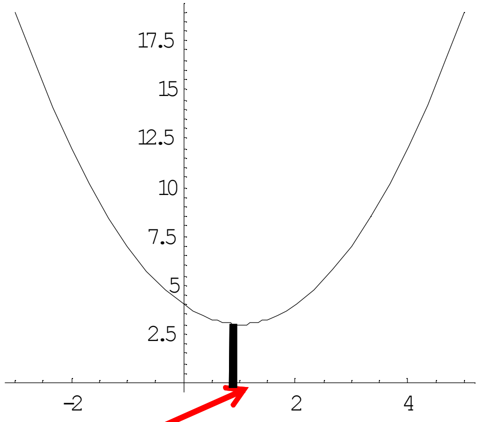
- **High-School algebra solution of the minimum problem.**

Denote $E(Y^2)=b, E(Y)=a,$

then

$$R(c) = c^2 - 2c \cdot a + b$$

is a **parabola** w.r.t. ***c*** with leading coefficient 1.



So the minimum of $R(c)$ is at

$$c_0 = a = E(Y).$$

© 2012 Mamikon Ginovyan

Best Predictor

- Now we show that the **best predictor** $\hat{Y} = g(X)$ depends on the **closeness measure**.

Theorem 2.

Assume that ***Y*** is a CRV with *pdf* ***f*(*y*)**, and **median *m***:

$$\int_{-\infty}^{m} f(y)dy = \int_{m}^{\infty} f(y)dy = \frac{1}{2}.$$

Then

$$\min_{c} E|Y - c| = E|Y - m|,$$

that is, in this case the **best predictor** $\hat{Y}$ is the **median** of RV ***Y*.**

.

© 2012 Mamikon Ginovyan

Theorem 2-Proof

Proof (Calculus).

Denote $R_1(c) = E|Y-c|$.

Since

$$|y-c| = \begin{cases} (y-c), & y \geq c \\ -(y-c), & y < c, \end{cases}$$

we have $R_1(c) = E|Y-c| = \int\limits_{-\infty}^{\infty} |y-c| f(y)dy$

$$= -\int\limits_{-\infty}^{c} (y-c)f(y)dy + \int\limits_{c}^{\infty} (y-c)f(y)dy.$$

Hence

$$\frac{dR_1(c)}{dc} = \int\limits_{-\infty}^{c} f(y)dy - \int\limits_{c}^{\infty} f(y)dy \overset{set}{=} 0$$

© 2012 Mamikon Ginovyan

Theorem 2-Proof

- The solution of this equation is

$$\boldsymbol{c = m = median}.$$

- This is a **minimum point** since

$$\frac{d^2 R_1(c)}{dc^2} = f(c) + f(c) = 2f(c) > 0.$$

- (by **Fundamental Theorem of Calculus**)

© 2012 Mamikon Ginovyan

Best MSE Predictor (General case)

- Now we use the definition and properties of **conditional expectations** to solve the **MSE-Prediction Problem in general case:**
- **Find** the **best MSE predictor** of a RV $\boldsymbol{Y}$ given a RV $\boldsymbol{X}$ or a random vector $\underline{X} = (X_1, \ldots, X_n)$, that is, find a function $g(\cdot)$ that **minimizes** the **Mean Square Error:**

$$MSE = E(Y - g(\underline{X}))^2 \rightarrow \min.$$

© 2012 Mamikon Ginovyan

Best MSE Predictor (General case)

Theorem 3.

If $\underline{X} = (X_1, \ldots, X_n)$ is any random vector and ***Y*** is any RV, then either

(a) $E(Y - g(\underline{X}))^2 = \infty$ for any function *g*, or

(b) $\min\limits_{g(x)} E(Y - g(\underline{X}))^2 = E(Y - E(Y \mid \underline{X}))^2,$

where $g(x)$ runs over all functions.

Thus, $\hat{Y} = g_0(\underline{X}) = E(Y \mid X)$

is the **unique** **best MSE predictor** of ***Y***.

© 2012 Mamikon Ginovyan

Theorem 3-Proof

- **Proof** of **Theorem 3**.

We have

$$E[Y-g(X)]^2 = E[(Y-E(Y|X))+(E(Y|X)-g(X))]^2$$

$$= E[Y-E(Y|X)]^2 + E[g(X)-E(Y|X)]^2$$

$$+2E\left[(Y-E(Y|X)(E(Y|X)-g(X))\right].$$

Conditioning expectation on $\boldsymbol{Y}$, that is, using the formula

$$E\left[X\right] = E\left[E\left[X \mid Y\right]\right],$$

it can be shown that the **last (cross) term is equal to zero**.

© 2012 Mamikon Ginovyan

Theorem 3-Proof

Thus,

$$(1) \quad E[Y-g(X)]^2 = E[Y-E(Y|X)]^2 + E[g(X)-E(Y|X)]^2$$

$$\geq E[Y-E(Y|X)]^2 \qquad \textit{for} \quad \textit{all} \quad g(\cdot).$$

The choice $g_0(X) = E(Y|X)$ will give equality.

© 2012 Mamikon Ginovyan

Best MSE Predictor

- The problem in finding the best MSE-Predictor is solved by **Theorem 3.**
- **Two difficulties** of the solution are:

(a)We need to know the **joint distribution** of ***X*** and ***Y*** in order to compute the best predictor

$$\hat{Y} = g(X) = E(Y \mid X).$$

(b) The best predictor (or equivalently the regression curve $g_0(x) = E(Y \mid X = x)$ of ***Y*** on ***X***) may be **complicated** function of ***x*** or **hard** to find.

© 2012 Mamikon Ginovyan

Best MSE-Linear Predictor

- We can **avoid both objections** by looking for a predictor which is best within a class of **simple (linear)** predictors.

Definition 1.

Any RV of the form $g(X) = a + bX$

is called a **linear predictor** and any such variable with $a = 0$ (i.e. $g(X) = bX$) is called a **zero intercept linear predictor**.

© 2012 Mamikon Ginovyan

Best MSE-Linear Predictor

Definition 2.

1. The **numbers** a_0 and b_0 for which the linear predictor $g_0 = a_0 + b_0 X$ minimizes

$$MSE = E(Y - \hat{Y})^2 = E[Y - (a + b X)]^2,$$

that is,

$$\min_{a,b} E[Y - (a + b X)]^2 = E[Y - (a_0 + b_0 X)]^2$$

are called the **regression intercept** (a_0) and **regression slope** (b_0) of ***Y*** on ***X*,** respectively.

© 2012 Mamikon Ginovyan

Best MSE-Linear Predictor

Definition 2.

2. The line $y = g_0(x) = a_0 + b_0 x$ is called the **regression line** of ***Y*** on ***X***.

3. The RV $\hat{Y} = g_0(X) = a_0 + b_0 X$ is the **best MSE-linear predictor** for ***Y*** given ***X***.

© 2012 Mamikon Ginovyan

Best MSE-Linear Predictor

- **How to find the best MSE-linear-predictor?**

The answer is given by the following theorem.

Theorem 4.

Suppose that $E(X^2)$ and $E(Y^2)$ are finite and ***X*** and ***Y*** are **not constant**. Then

(a-1) The **unique best zero intercept MSE-linear predictor** is given by

$$\hat{Y} = g_0(X) = b_0 X \quad \text{with} \quad b_0 = \frac{E[XY]}{E[X^2]}.$$

© 2012 Mamikon Ginovyan

Theorem 4.

(a-2) The **MSE- prediction error** is given by

$$E[Y - b_0 X]^2 = \frac{E(X^2)E(Y^2) - (E[XY])^2}{E(X^2)}$$

$$= E[Y^2] - \frac{(E[XY])^2}{E[X^2]}.$$

(b-1) The **unique best MSE-linear predictor** is given by

$$\hat{Y} = g_0(X) = a_0 + b_0 X$$

with

(1) $$b_0 = \frac{Cov(X,Y)}{s^2(X)} = r(X,Y)\frac{s(Y)}{s(X)},$$

(2) $$a_0 = E(Y) - b_0 E(X) = E(Y) - r(X,Y)\frac{s(Y)}{s(X)}E(X),$$

© 2012 Mamikon Ginovyan

Theorem 4.

where $s^{2}(X) = Var(X), \quad s(X) = \sqrt{Var(X)}$

and $\quad r(X,Y) = \dfrac{Cov(X,Y)}{s(X)s(Y)}.$

(b-2) The **regression line** of ***Y*** on ***X*** is given by

(3) $y = a_0 + b_0 x \Leftrightarrow \dfrac{y - E(Y)}{s(Y)} = r(X,Y)\dfrac{x - E(X)}{s(X)}.$

(b-3) The **MSE - prediction error** is given by

(4) $E[Y - \hat{Y}]^2 = E[Y - (a_0 + b_0 X)]^2 = [1 - r^{2}(X,Y)]s^{2}(Y).$

© 2012 Mamikon Ginovyan

Proofs - Preliminaries

The quadratic function

$$y = f(x) = ax^2 + bx + c, a \neq 0,$$

a, b and c are real constants.

Standard form: $$y = a\left[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a^2}\right].$$

Extremum point: $$y' = 2ax + b/ = 0$$

$$\Rightarrow x_0 = -\frac{b}{2a}.$$

© 2012 Mamikon Ginovyan

The quadratic function

Minimum and Maximum values:

- If $a > 0$ (Fig. 1)

$$y_0 = \min_{x(a>0)} f(x) = f\left(-\frac{b}{2a}\right) = -\frac{b^2 - 4ac}{4a} = \frac{4ac - b^2}{4a}.$$

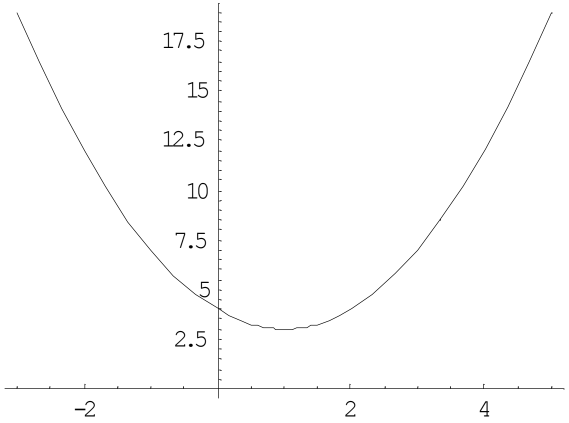


Fig.1

© 2012 Mamikon Ginovyan

The quadratic function

Minimum and Maximum values:

- If $a < 0$ (Fig. 2)

$$y_0 = \max_{x(a<0)} f(x) = f\left(-\frac{b}{2a}\right) = -\frac{b^2 - 4ac}{4a} = \frac{4ac - b^2}{4a}.$$

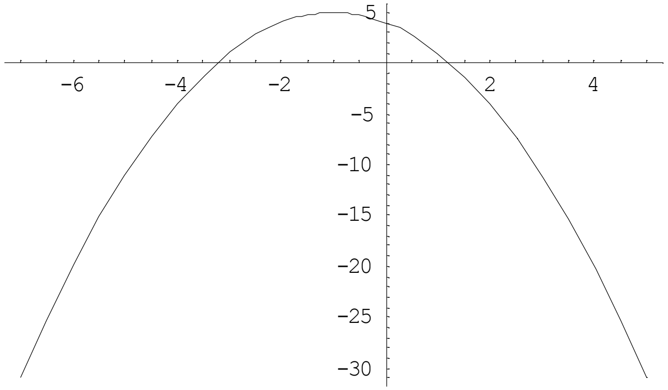


Fig.2

© 2012 Mamikon Ginovyan

Theorem 4.-Proof

Proof (a).

Let $g(X) = b\,X$ be a zero intercept linear predictor.

We expand $E[Y - b\,X]^2$

to get

$$E[Y - b\,X]^2 = E(Y^2) - 2b\,E(XY) + b^2 E(X^2) = ab^2 - 2bb + c,$$

where $a = E(X^2),\ b = E(XY),\ c = E(Y^2)$

This is a quadratic function w.r.t. b with leading coefficient

$$a = E(X^2) > 0.$$

© 2012 Mamikon Ginovyan

Theorem 4.-Proof

Therefore $E[Y - b\,X]^2$ is uniquely minimized by

(5) $$b_0 = -\frac{(-2b)}{2a} = \frac{b}{a} = \frac{E(XY)}{E(X^2)},$$

and the minimum (the **mean squared prediction error**) is

(6) $$E[Y - b_0 X]^2 = \frac{ac - b^2}{a} = \frac{E(X^2)E(Y^2) - (E[XY])^2}{E(X^2)}.$$

© 2012 Mamikon Ginovyan

Theorem 4.-Proof

Proof (b). Using the identity (see proof of Th.1)

$$(7) \qquad E(Z-c)^2 = Var(Z) + [E(Z)-c]^2$$

with $Z = Y - bX$ and $c = a$,

we obtain

$$E[Y-(a+bX)]^2 = E[(Y-bX)-a]^2$$

$$= Var(Y-bX) + [E(Y) - b\,E(X) - a]^2.$$

© 2012 Mamikon Ginovyan

Theorem 4.-Proof

Since both terms on the right are non-negative, whatever b, the quantity

$$E[Y-(a+bX)]^2$$

is uniquely minimized by taking

$$a = E(Y) - b\,E(X). \tag{8}$$

Substituting this value of a into $E[Y-(a+bX)]^2$, we see that b we seek minimizes

$$E[Y-(a+bX)]^2 = E([Y-E(Y)] - b[X-E(X)])^2 \tag{9}$$

$$= E(Y_1 - b\,X_1)^2.$$

© 2012 Mamikon Ginovyan

Theorem 4.-Proof

Now we can apply the result in part (a) on **zero intercept** linear predictors to the RV's

(10) $X_1 = X - E(X)$ and $Y_1 = Y - E(Y)$

to conclude that the number

(11) $$b_0 = \frac{E[X_1 Y_1]}{E[X_1^2]} = \frac{E(X - EX)(Y - EY)}{E(X - EX)^2} = \frac{Cov(X, Y)}{s^2(X)}$$

is the unique minimizing value.

Thus, formula (1) is proved.

© 2012 Mamikon Ginovyan

Theorem 4.-Proof

- **To prove (2),** we substitute b_0 from (11) into (8).

- **To Prove (4),** we apply (6) to X_1 and Y_1 defined by (10)

$$E[Y-(a_0+b_0X)]^2 = E[Y_1 - b_0X_1]^2$$

$$= \frac{E(X_1^{\;2})E(Y_1^{\;2}) - (E[X_1Y_1])^2}{E(X_1^{\;2})}$$

$$= \frac{E(X-EX)^2 E(Y-EY)^2 - [E(X-EX)(Y-EY)]^2}{E(X-EX)^2}$$

© 2012 Mamikon Ginovyan

Theorem 4.-Proof

$$= \frac{s^2(X)s^2(Y) - [Cov(X,Y)]^2}{s^2(X)}$$

$$= \frac{s^2(X)s^2(Y) - r^2(X,Y)s^2(X)s^2(Y)}{s^2(X)}$$

$$= \frac{s^2(X)s^2(y)[1 - r^2(X,y)]}{s^2(X)}$$

$$= [1 - r^2(X,y)]s^2(y).$$

- This completes the proof of Theorem 4.

An Example

v **Example.**

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two RV's such that

$$s^{2}(Y)=10 \quad \text{and} \quad r(X,Y)=.5.$$

Then

1. If we **ignore** $\boldsymbol{X}$ and predict $\boldsymbol{Y}$ as simply **E(*Y*)**:

$$\hat{Y}=E(Y),$$

we will have a **MSE - prediction error** equal to the variance of $\boldsymbol{Y}$, namely 10:

$$E(Y-\hat{Y})^{2}=E(Y-EY)^{2}=s^{2}(Y)=10.$$

© 2012 Mamikon Ginovyan

An Example

2. If we use the **regression line** of ***Y*** on ***X*** to predict ***Y:***

$$\hat{Y} = a_0 + b_0 X,$$

then for the **MSE - prediction error** we will have (see formula (4)):

$$E[Y - \hat{Y}]^2 = E[Y - (a_0 + b_0 X)]^2$$

(4`) $$= [1 - r^2(X, Y)]s^2(Y)$$

$$= (1 - .25)(10) = (.75)(10) = 7.5,$$

that is, a **25%** **reduction** comparing with **case 1.**

© 2012 Mamikon Ginovyan

Residual Variance

For this reason, it is often said that

"the square of the correlation coefficient $= r^2(X,Y)$ **is the proportion of the variance of** $\boldsymbol{Y}$ **accounted for by linear regression on** $\boldsymbol{X}$**",**

and the **MSE - prediction error**

$$[1 - r^2(X,Y)]s^2(Y)$$

is called the **residual variance** (after **linear regression** on $\boldsymbol{X}$).

© 2012 Mamikon Ginovyan

Correlation Inequality

Corollary (Correlation Inequality).

For any two RV's ***X*** and ***Y*** such that

$$s^2(X) < \infty \quad and \quad s^2(Y) < \infty,$$

(a) $|r(X,Y)| \le 1$

(b) $|r(X,Y)| = 1$ **if and only if**

1) ***X*** or ***Y*** is a constant, or
2) ***X*** and ***Y*** are **linearly related**, more precisely:

$$Y - E(Y) = \frac{Cov(X,Y)}{s^2(Y)}[X - E(X)].$$

Proof: Follows from Corollary 1, applying to the RV's

$$X_1 = X - EX \quad and \quad Y_1 = Y - EY.$$

© 2012 Mamikon Ginovyan

Correlation Coefficient

Remark 1.

The **square** of correlation coefficient, $r^2(X, Y)$, or the **absolute value** $|r(X, Y)|$ can be regarded as a **measure of the utility** of using $\boldsymbol{X}$ in a linear manner to predict $\boldsymbol{Y}$.

The correlation coefficient $r(X, Y)$ measures (roughly), the **amount** and **sign** of **linear relationship** between the RV's $\boldsymbol{X}$ and $\boldsymbol{Y}$.

More precisely:

© 2012 Mamikon Ginovyan

Correlation Coefficient

1) If $r(X,Y)=1$, then $Y=a+bX, b>0,$

   high utility (accurate prediction)

2) If $r(X,Y)=-1$, then $Y=a+bX, b<0,$

   high utility (accurate prediction)

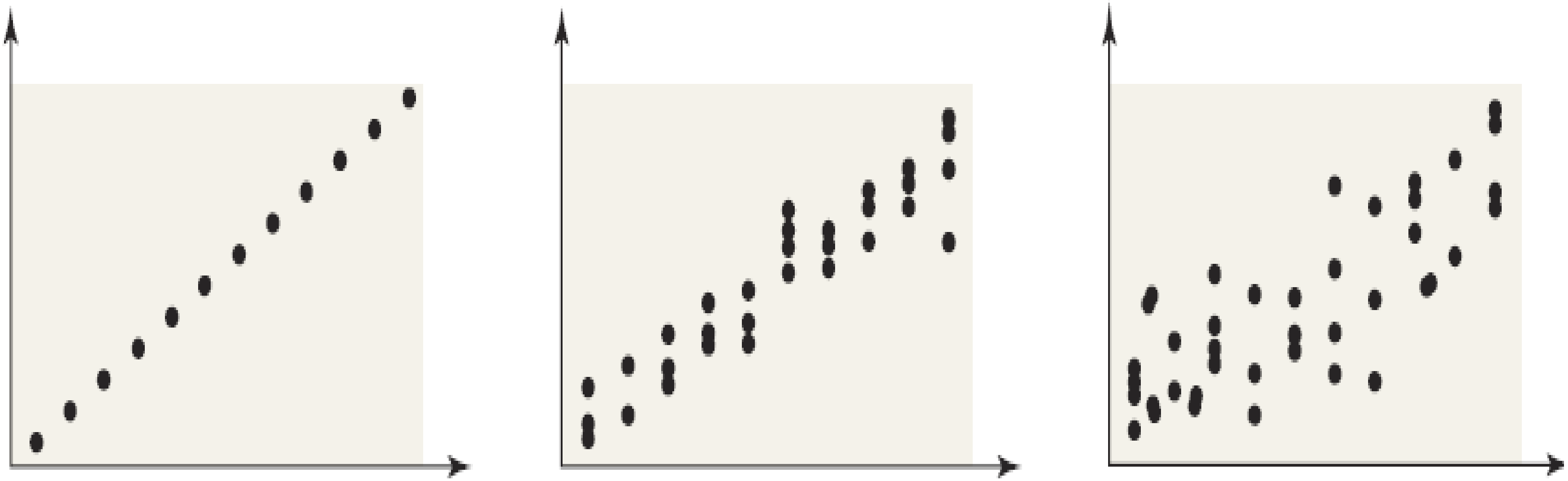
3) If $r(X,Y)=0$, then $\boldsymbol{X}$ and $\boldsymbol{Y}$ are uncorrelated,

   low utility (inaccurate prediction).

© 2012 Mamikon Ginovyan

Correlation Coefficient

Positive Correlation



(a) Perfect positive linear relation, $r = 1$

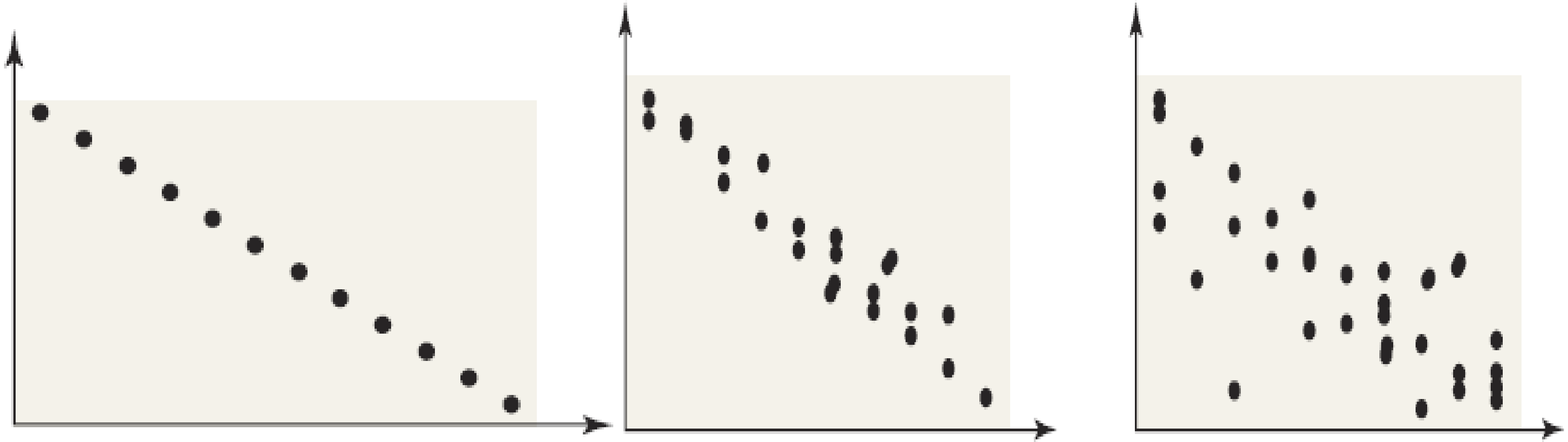
(b) Strong positive linear relation, $r \approx 0.9$

(c) Moderate positive linear relation, $r \approx 0.4$

high utility (accurate prediction)

Correlation Coefficient

Negative Correlation



(d) Perfect negative linear relation, $r = -1$

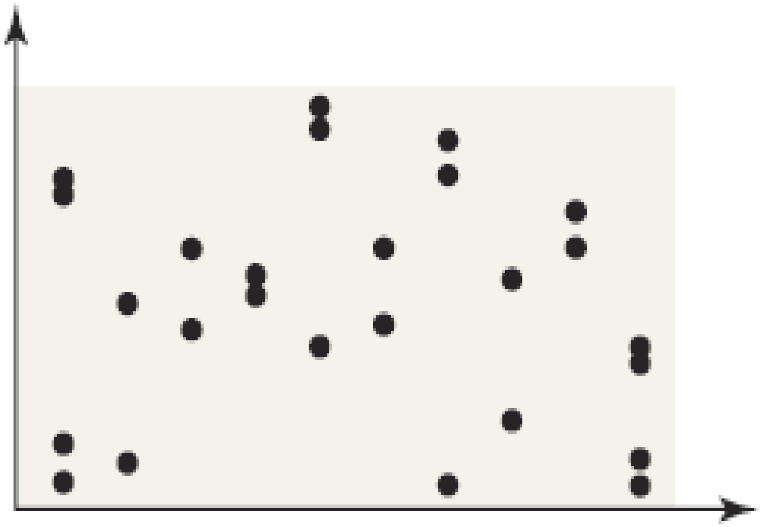
(e) Strong negative linear relation, $r \approx -0.9$

(f) Moderate negative linear relation, $r \approx -0.4$

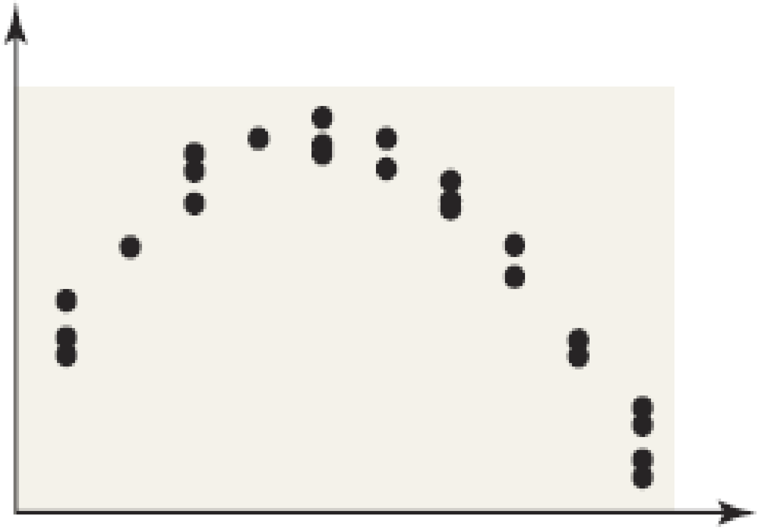
high utility (accurate prediction)

Correlation Coefficient

No Correlation



(g) No linear relation, r close to 0.



(h) No linear relation, r close to 0.

low utility (inaccurate prediction)

Best Predictor vs. Best Linear Predictor

Remark 2.

Let $\hat{Y} = E[Y|X]$ be the **best predictor** ,and

$\hat{Y}_L = a_0 + b_0 X$ be the **best linear predictor**.

If the best predictor $\hat{Y} = E[Y|X]$ is of the form $\hat{Y} = a + b X$, then $a = a_0, b = b_0$, since, if **the best predictor is linear**, it must **coincide** with the best linear predictor. (See **Example 1** below).

- **In general, the best predictor and the best linear predictor differ** (see **Example 2** below).

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

v **Example 1.**

Suppose that X and Y have a bivariate normal distribution:

$$(X,Y) : N(m_1, m_2, s_1^2, s_2^2, r).$$

a) Find the **best predictor** of Y using X, that is, the **regression curve** of Y on X, and show that it **coincides** with the **best linear predictor**.

b) Find the MSE-prediction error of the best predictor.

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

Solution.

Recall that a two-dimensional random vector (***X, Y***) has a **bivariate normal distribution**

$$(X,Y) : N(m_1, m_2, s_1^{\ 2}, s_2^{\ 2}, r)$$

if its *pdf* *f(x)* is given by

$$(1) \qquad f(x,y) = \frac{1}{2ps_1s_2\sqrt{1-r^2}} \times$$

$$\exp\left\{-\frac{1}{2(1-r^{2)}}\left[\left(\frac{x-m_1}{s_1}\right)^2 - 2r\left(\frac{x-m_1}{s_1}\right)\left(\frac{y-m_2}{s_2}\right)+\left(\frac{y-m_2}{s_2}\right)^2\right]\right\},$$

Best Predictor vs. Best Linear Predictor

where

$$m_1 = E(X), \quad m_2 = E(Y), \quad s_1^{\,2} = Var(X), \quad s_2^{\,2} = Var(Y),$$

$$r = r(X,Y) = Cor(X,Y) = \frac{Cov(X,Y)}{s_1 s_2}.$$

Observe that

1. $X \sim N(m_1, s_1^{\,2})$ and $Y \sim N(m_2, s_2^{\,2})$

2. If $r = 0,$ then

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

that is, ***X*** and ***Y*** are **independent**.

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

- **Conclusion.**

If $\boldsymbol{X}$ and $\boldsymbol{Y}$ have a bivariate normal distribution, then $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent if and only if they are uncorrelated $(r = 0)$.

Solution (a).

To compute the best predictor $\hat{Y}$ of $\boldsymbol{Y}$ using $\boldsymbol{X}$, which is the conditional expectation

$$\hat{Y} = E(Y|X)$$

we first compute the **conditional pdf** $f(y|x)$:

$$f(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}. \tag{2}$$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

Since $X \sim N(m_1, s_1^{\,2})$, we have

$$(3) \qquad f_X(x) = \frac{1}{\sqrt{2p}\, s_1} e^{-\frac{(x-m_1)^2}{2s_1^{\,2}}}$$

Substituting (1) and (3) into (2) we obtain

$$f(y|x) = \frac{1}{s_2\sqrt{2p(1-r^2)}} \times$$

$$\exp\left\{-\frac{1}{2(1-r^2)}\left[\left[1-(1-r^{2)}\right]\frac{(x-m_1)^2}{s_1^{\,2}} - \frac{2r}{s_1 s_2}(x-m_1)(y-m_2) + \frac{(y-m_2)^2}{s_2^{\,2}}\right]\right]\right\}$$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

$$= \frac{1}{s_2\sqrt{2p(1-r^2)}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(y-m_2)^2}{s_2} - r\,\frac{(x-m_{1)}}{s_1}\right]^2\right\}$$

$$= \frac{1}{s_2\sqrt{2p(1-r^2)}}\exp\left\{-\frac{1}{2s_2^2(1-r^2)}\left[y-\left[m_2+\frac{r s_2}{s_1}(x-m_1)\right]\right]^2\right\}.$$

- **Thus,** the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{X}=x$ is **normal** $N(m, s^2)$, where

$$m = m_2 + \frac{r s_2}{s_1}(x-m_1)$$

$$s^2 = s_2^2(1-r^2).$$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

Since for $X \sim N(m, s^2)$,

$$m = E(X) \quad \text{and} \quad s^2 = Var(X),$$

we conclude that the **best predictor** of ***Y*** given ***X*** is the **linear function**

$$\hat{Y} = E[Y | X] = m_2 + \frac{r s_2}{s_1}(X - m_1).$$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

Solution (b). Since

$$E\left[\left(Y - E(Y \mid X = x)\right)^2 \middle| X = x\right] = s_2^2(1 - r^2)$$

is independent of ***x***, the MSE of the best predictor is

$$E(Y - \hat{Y})^2 = E(Y - E(Y \mid X))^2 = s_2^2(1 - r^2).$$

Remark-Problem. **Similarly can be found** the corresponding formulas for

$$\hat{X} = E(X \mid Y).$$

Best Predictor vs. Best Linear Predictor

- **Example 2.**

Suppose the DRV's ***X*** and ***Y*** have the following joint probability distribution *f(x, y)* :

X \ Y	**0**	**1**	**2**	**3**	$f_X(x)$
1/4	.1	.05	.05	.05	.25
1/2	.025	.025	.1	.1	.25
1	.025	.025	.1	.3	.5
$f_Y(y)$	.15	.1	.3	.45	1

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

a) Find the **best MSE predictor** $\hat{Y} = E[Y|X]$ of RV ***Y*** given ***X***.

b) Find the **MSE of the best predictor**: $s^2 = E[Y - E(Y|X)]^2$.

c) Find the **best linear MSE predictor** $\hat{Y}_L = a_0 + b_0 X$ of ***Y*** given ***X***.

d) Find the **MSE of the best linear predictor:** $s_L^2 = E[Y - \hat{Y}_L]^2$.

e) Find the **ratio** $\dfrac{s_L^2}{s^2}$ and state your conclusion.

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

Solution.

a) We have

$$E[Y|X=1] = \sum_{k=0}^{3} kP[Y=k|X=1]$$

$$= \sum_{k=0}^{3} k \frac{P(Y=k, X=1)}{P(X=1)}$$

$$= \frac{1}{P(X=1)} \sum_{k=0}^{3} kf(k,1)$$

$$= \frac{1}{.5}[0(.025)+1(.025)+2(.15)+3(.3)] = 2.45.$$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

Similarly we find

$$E\left[Y \middle| X = \frac{1}{2}\right] = 2.1 \qquad and \qquad E\left[Y \middle| X = \frac{1}{4}\right] = 1.2.$$

b) $s^{2} = E[Y - E(Y|X)]^{2}$

$$= \sum_{i=1}^{3}\sum_{j=1}^{4}[y_i - E(Y|X = x_i)]^2 f(x_i, y_i)$$

$$= (1.2)^2(.1) + (1 - 1.2)^2(.05) + \cdots + (3 - 2.45)(.3) = .885.$$

So $s^{2} = .885.$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

c) $\hat{Y}_L = a_0 + b_0 X,$ find a_0 *and* b_0.

First we find $b_0 = \dfrac{Cov(X,Y)}{Var(X)}$.

(1) $E[X] = \frac{1}{4}(.25) + \frac{1}{2}(.25) + 1(.5) = .6875 \approx .69.$

(2) $E[Y] = 0(.15) + 1(.1) + 2(.3) + 3(.45) = 2.05.$

(3) $E[X^2] = (\frac{1}{4})^2(.25) + (\frac{1}{2})^2(.25) + 1^2(.5) \approx .578.$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

(4) $$E[XY] = \sum_{i=1}^{3}\sum_{j=1}^{4} x_i y_j f(x_i, y_j)$$

$$= \frac{1}{4}[0(.1) + 1(.05) + 2(.05) + 3(.05)]$$

$$+ \frac{1}{2}[0(.025) + 1(.025) + 2(.1) + 3(.1)]$$

$$+ 1[0(.025) + 1(.025) + 2(.15) + 3(.3)] = 1.5625.$$

So, $$Cov(X,Y) = E[XY] - E[X]E[Y] = 1.56 - 1.4 = .16.$$

$$Var(X) = E[X^2] - (E[X])^2 = .11.$$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

Thus, $$b_0 = \frac{Cov(X,Y)}{Var(X)} = \frac{.16}{.11} = 1.45.$$

Then $a_0 = E[Y] - b_0 E[X] = 2.05 - (1.45)(.69) = 1.05.$

Therefore, $\hat{Y}_L = a_0 + b_0 X = 1.05 + 1.45X.$

d) $$s_L^2 = E[Y - \hat{Y}_L]^2 = E[Y - 1.45X - 1.05]^2$$

$$= E[Y^2] + (1.45)^2 E[X^2] + (1.05)^2 \qquad (5)$$

$$-(2.9)E[XY] - (2.1)E[Y] + (3.045)E(X).$$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

Using (1) - (4) and taking into account that

$$E[Y^2] = 0(.15) + 1(.1) + 4(.3) + 9(.45) = 5.35,$$

from (5) we obtain $s_L^2 = .93.$

e) For the ratio $\dfrac{s_L^2}{s^2}$ we have

$$\frac{s_L^2}{s^2} = \frac{E[Y - \hat{Y}_L]^2}{E[Y - E(Y|X)]^2} = \frac{.93}{.885} = 1.05. \qquad (6)$$

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor

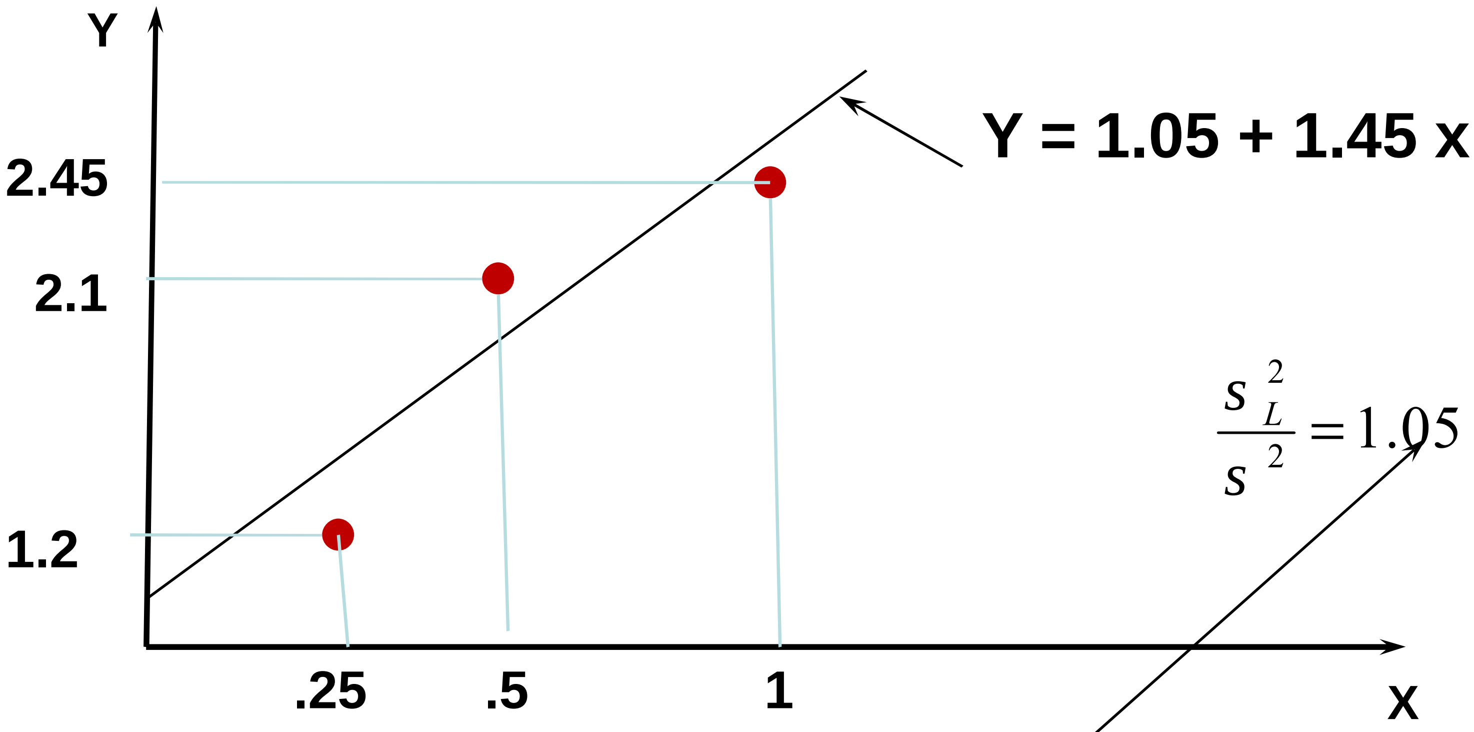
- **Conclusion.**

In this example, the best liner predictor $\hat{Y}_L$ and the best predictor $\hat{Y}$ **differ**.

	$\hat{Y}$	$\hat{Y}_L$
X = 1/4	1.2	1.41
X = 1/2	2.1	1.775
X = 1	2.45	2.5

© 2012 Mamikon Ginovyan

Best Predictor vs. Best Linear Predictor



$$\frac{s_L^2}{s^2} = 1.05$$

The three dots give the **best predictor** ($\hat{Y}$). The line **Y = 1.05 + 1.45x** represents the **best linear predictor.** A loss about **5%** reflected in (6), is incurred by using the best linear predictor $\hat{Y}_L$ **= 1.05 + 1.45x.**

A Surprising Example

- **Two RV's that are uncorrelated even though one of them may be predicted perfectly from the other.**

- **Motivation.**

The point of this example is to further expose the **fallacy** that **uncorrelated RV's are independent**.

Recall that, we showed

(a) if X and Y are independent, they also are uncorrelated,

(b) the converse, generally, is not true.

© 2012 Mamikon Ginovyan

A Surprising Example

(c) If X and Y are **bivariate normal** RV's, then **the converse is true** (notice that if **only marginals are normal**, again, the **converse, generally, is not true**).

- **In fact,** we will show that uncorrelated RV's may be directly related by a functional relationship and, hence, may be dependent.
- The example will show that the **covariance** (or **correlation**) **strictly provides a measure of linear dependence** between RV's, and may **not be sensitive to nonlinearities**.

© 2012 Mamikon Ginovyan

A Surprising Example

- **A Surprising Example.**

Let $\boldsymbol{X} \sim \boldsymbol{U}(-a, a)$ with some $a>0$. Then $m = E[X] = 0$, and

$$m'_{2k+1} = E[X-m]^{2k+1} = E[X]^{2k+1} = 0 \quad \text{for all } k = 1,2,\ldots.$$

Consider the RV $Y = X^2$.

It is clear that $\boldsymbol{X}$ and $Y = X^2$ are **dependent**, at the same time for covariance we have

$$Cov(X,Y) = Cov(X,X^2) = E[X \cdot X^2] - E[X]E[X^2]$$
$$= E[X^3] - E[X]E[X^2] = 0 \qquad (4)$$

that is, the RV's X and $Y = X^2$ are **uncorrelated** **(but dependent)**.

© 2012 Mamikon Ginovyan

A Surprising Example

- **Remark 1.**

  The example seems **especially surprising** because there is direct functional relationship (**dependence**) between $\boldsymbol{X}$ and $Y = X^2$.

  Nevertheless, the **best linear predictor** of $Y = X^2$ based on $\boldsymbol{X}$ is **constant**, that is, $Y = X^2$ can be **predicted perfectly** from $\boldsymbol{X}$.

© 2012 Mamikon Ginovyan