JTMS-12: Probability and Random Processes

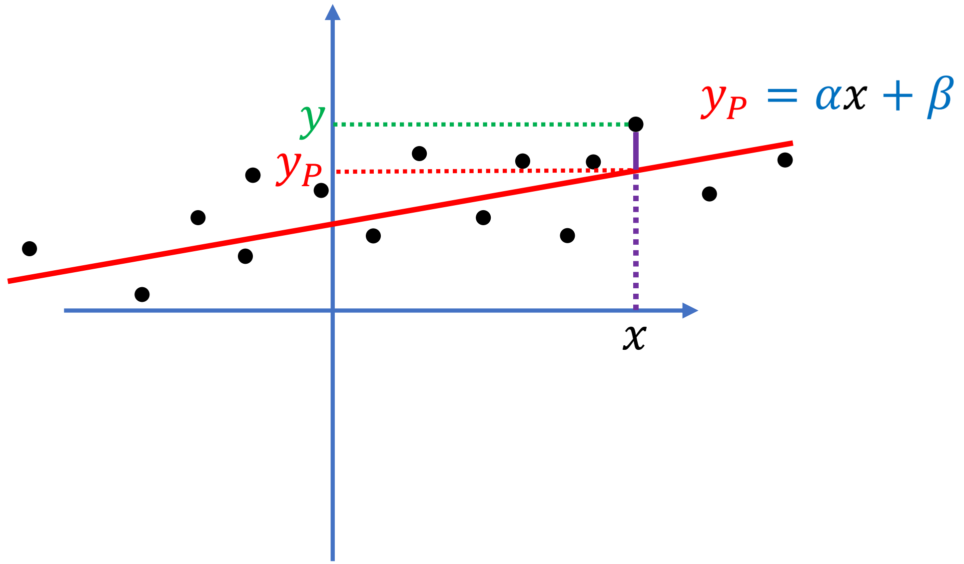
Fall 2020

M. Bode

Lecture 17

Recap

Moments – Joint Moments – Linear Prediction



Consider two r.v.s X and Y. Based on X, predict the Y …

Inspired by previous observations, we set up a linear model,

$$Y_P = \alpha X + \beta$$

… and try to predict with minimal quadratic error… such that

$$\varepsilon^2 = E[(Y - Y_P)^2] = E[(Y - \alpha X - \beta)^2] \rightarrow min.$$

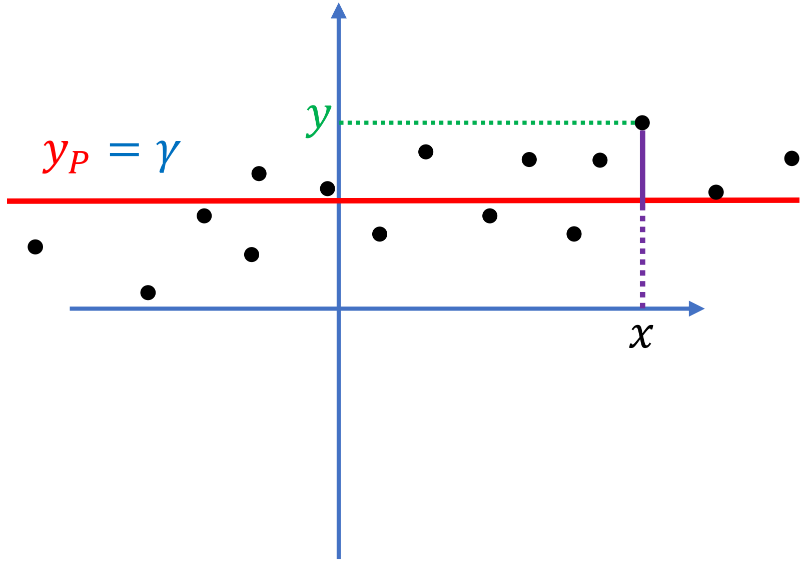
We require

$$0 = \frac{\partial}{\partial \alpha}\varepsilon^2 \;\; and \;\; 0 = \frac{\partial}{\partial \beta}\varepsilon^2$$

Result:

$$\boxed{\alpha = \frac{Cov[X,Y]}{\sigma_X^2}} \qquad \boxed{\beta = E[Y] - \frac{Cov[X,Y]}{\sigma_X^2}E[X]}$$

Lecture 17



Is this a good result?

How to compare?

Here is a simpler model:

$$Y_P = \gamma \quad (gamma)$$

Minimize the quadratic error:

$$\varepsilon^2 = E[(Y - Y_P)^2] = E[(Y - \gamma)^2] \to min.$$

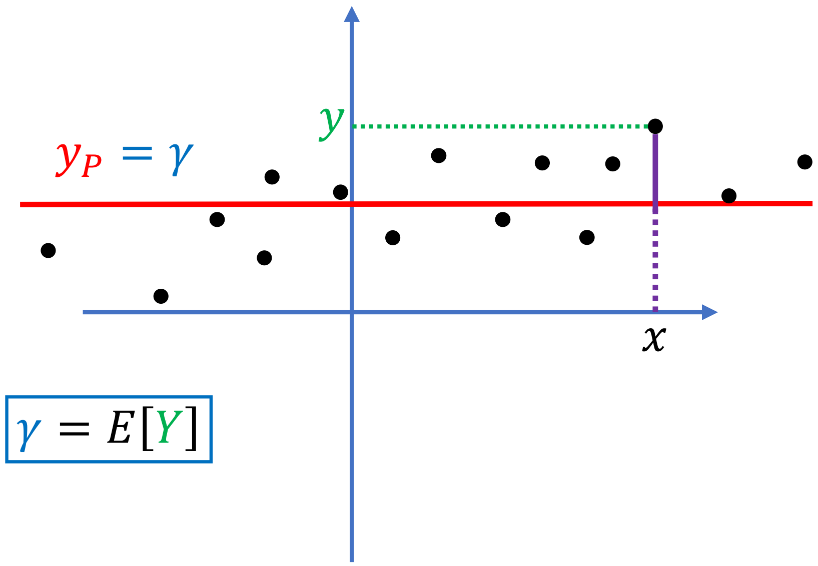
$$\Rightarrow 0 = -2E[Y - \gamma]$$

$$\Rightarrow \boxed{\gamma = E[Y]}$$

This one should be worse …

… larger mean square error $\varepsilon^2 = E[(Y - Y_P)^2]$ it seems.

Lecture 17



$\boxed{\gamma = E[Y]}$

Compare the resulting values of the mean square error

$$\varepsilon^2 = E[(Y - Y_P)^2]$$

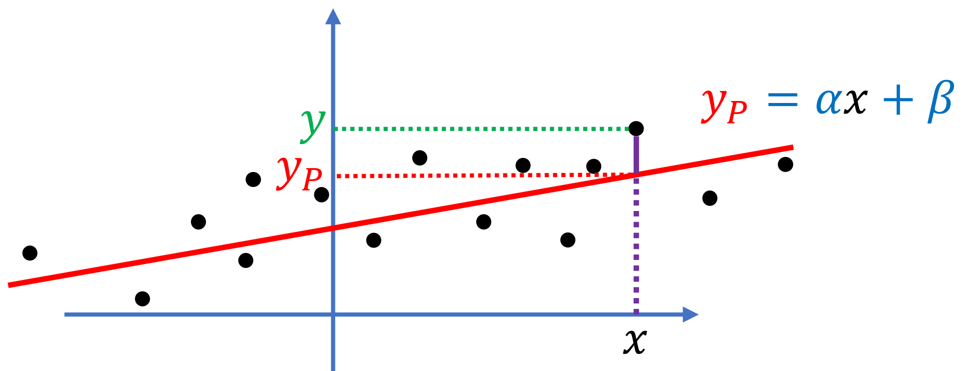
2nd model:

$$\varepsilon_2^2 = E[(Y - \gamma)^2] = E[(Y - E[Y])^2]$$

$$\Rightarrow \boxed{\varepsilon_2^2 = Var[Y]}$$

Model 2 is correct on average … but cannot cover the fluctuations.

Lecture 17



$$\boxed{\alpha = \frac{Cov[X,Y]}{\sigma_X^2}} \qquad \boxed{\beta = E[Y] - \frac{Cov[X,Y]}{\sigma_X^2}E[X]}$$

Compare the resulting values of the mean square error

$$\varepsilon^2 = E[(Y - Y_P)^2]$$

1st model:

$$\varepsilon_1^2 = E[(Y - \alpha X - \beta)^2]$$

$$= E\left[\left(Y - \frac{Cov[X,Y]}{\sigma_X^2}X - E[Y] + \frac{Cov[X,Y]}{\sigma_X^2}E[X]\right)^2\right]$$

$$= E\left[\left(Y - E[Y] - \frac{Cov[X,Y]}{\sigma_X^2}X + \frac{Cov[X,Y]}{\sigma_X^2}E[X]\right)^2\right]$$

$$= E[(Y - E[Y])^2] + E\left[\left(\frac{Cov[X,Y]}{\sigma_X^2}X - \frac{Cov[X,Y]}{\sigma_X^2}E[X]\right)^2\right]$$

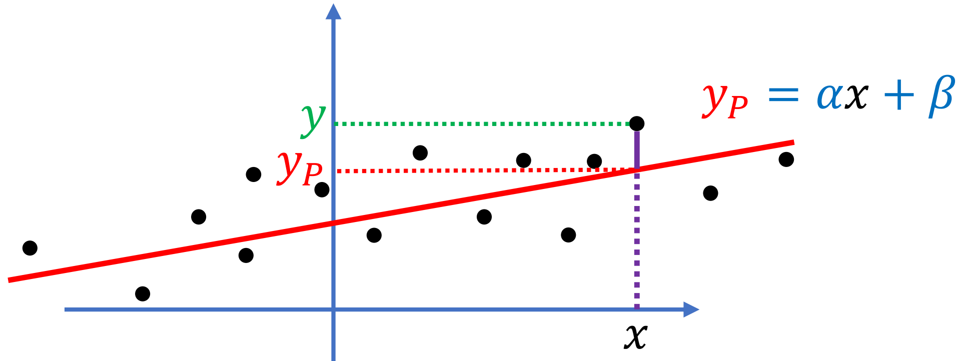
$$-2E\left[(Y - E[Y])\left(\frac{Cov[X,Y]}{\sigma_X^2}X - \frac{Cov[X,Y]}{\sigma_X^2}E[X]\right)\right]$$

$$= Var[Y] + \left(\frac{Cov[X,Y]}{\sigma_X^2}\right)^2 Var[X] - 2\frac{Cov[X,Y]}{\sigma_X^2}Cov[Y,X]$$

$$\Rightarrow \boxed{\varepsilon_1^2 = Var[Y] - \frac{Cov[X,Y]^2}{\sigma_X^2}}$$

$0 \leq \varepsilon_1^2 \leq \varepsilon_2^2$ … That is, model 1 is better …
it does cover the mean and part of the fluctuations.

Lecture 17



$$\boxed{\alpha = \frac{Cov[X,Y]}{\sigma_X^2}}$$

$$\boxed{\beta = E[Y] - \frac{Cov[X,Y]}{\sigma_X^2} E[X]}$$

$$\Rightarrow Var[Y] = Var[Y - Y_P] + Var[Y_P]$$

$$= E[(Y - Y_P)^2] + Var[Y_P]$$

$$= \varepsilon^2 + Var[Y_P] = \varepsilon^2 + \rho^2 Var[Y]$$

Usually, this explains only part of the observed fluctuations of Y.

How much? And how much remains unexplained?

$$Var[Y] = Var[Y - Y_P + Y_P]$$

$$= Var[Y - Y_P] + Var[Y_P] + 2Cov[Y - Y_P, Y_P]$$

Claim:

$$Cov[Y - Y_P, Y_P] = Cov[Y, Y_P] - Var[Y_P] = 0$$

Study:

$$Cov[Y, Y_P] = Cov\left[Y, \frac{Cov[X,Y]}{\sigma_X^2} X + \beta\right]$$

$$= \frac{Cov[X,Y]}{\sigma_X^2} Cov[Y, X] = \frac{Cov[X,Y]^2}{\sigma_X^2}$$

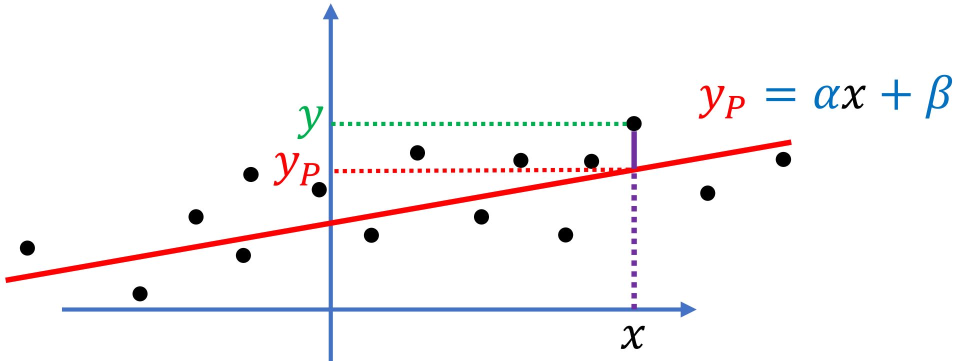
Also,

$$Var[Y_P] = Var\left[\frac{Cov[X,Y]}{\sigma_X^2} X + \beta\right] = \frac{Cov[X,Y]^2}{\sigma_X^2} \quad \blacksquare$$

Use correlation ρ:

$$= \frac{Cov[X,Y]^2}{\sigma_X^2 \sigma_Y^2} Var[Y] = \rho^2 Var[Y]$$

Lecture 17



$$\alpha = \frac{Cov[X, Y]}{\sigma_X^2}$$

$$\beta = E[Y] - \frac{Cov[X, Y]}{\sigma_X^2} E[X]$$

We found:

$$Var[Y] = Var[Y - Y_P] + Var[Y_P]$$

$$= \varepsilon^2 + Var[Y_P] = \varepsilon^2 + \rho^2 Var[Y]$$

Interpret via the correlation

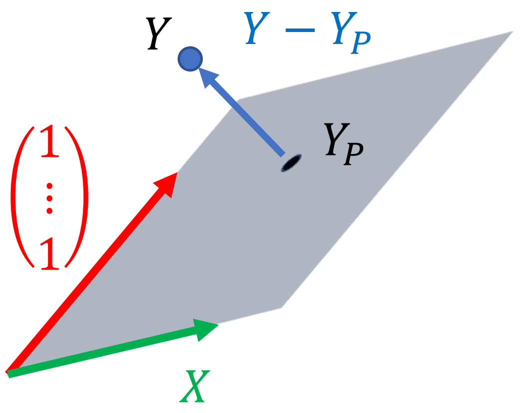
$$\rho = \frac{Cov[X, Y]}{\sigma_X \sigma_Y}$$

Original variance: $Var[Y]$

Explained variance: $Var[Y_P] = \rho^2 Var[Y]$

Unexplained variance: $\varepsilon^2 = (1 - \rho^2) Var[Y]$

Lecture 17



Model:

$$Y_P = \alpha X + \beta \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$$

Common wording:

The error is orthogonal to the data.

But use with care!

Different Perspective …

Observed data vectors X and Y are points in an n-dimensional space.

Task: For a given X, approximate Y based on a 2-dim. model

$$Y_P = \alpha X + \beta \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$$

The model describes points Y_P in a 2-dimensional linear subspace.

Interpretation: Within the model-subspace, find the point Y_P closest to Y … minimizing the distance $\|Y - Y_P\|$.

Solution: Draw a line orthogonal to the model-plane through Y. This line intersects the model-plane at the optimal point Y_P .

➔ $Y - Y_P \perp X$ and $Y - Y_P \perp \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$

Lecture 17

Chapter 4: **Expectation and Introduction to Estimation**

Chebyshev Inequalities

Full calculations can be hard … or impossible.

➔Approximations and bounds are useful tools.

$$\sigma_X^2 = Var[X] = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx$$

$$\geq \int_{|x-\mu_X|\geq\delta} (x-\mu_X)^2 f_X(x)dx \geq \delta^2 \int_{|x-\mu_X|\geq\delta} f_X(x)dx$$

$$= \delta^2 P[|X-\mu_X| \geq \delta]$$

$$\Rightarrow \boxed{P[|X-\mu_X| \geq \delta] \leq \frac{\sigma_X^2}{\delta^2}}$$

Probability of the tails is *small.*

Similar relations can be obtained for other even moments.

Lecture 17

Chapter 4: **Expectation and Introduction to Estimation**

Chebyshev Inequalities

$$\boxed{P[|X - \mu_X| \geq \delta] \leq \frac{\sigma_X^2}{\delta^2}}$$

Probability of the tails is *small.*

Important application:

Consider i.i.d. random variables $X_1, \ldots, X_n$ with finite but unknown mean μ_X and variance σ_X^2 and their average

$$\mathrm{Z} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

In order to estimate the mean μ_X, people try to use $\hat{\mu}_X = Z$.

Example:

Measure the same quantity many times, then take the average (noisy signals, voters, …)

Does that make sense?

Lecture 17

Chapter 4: **Expectation and Introduction to Estimation**

Chebyshev Inequalities

$$\boxed{P[|X-\mu_X| \geq \delta] \leq \frac{\sigma_X^2}{\delta^2}}$$

Probability of the tails is *small*.

1)

$$\mu_Z = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu_X$$

2)

$$\sigma_Z^2 = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} n\sigma_X^2 = \frac{\sigma_X^2}{n}$$

$$\Rightarrow P[|Z-\boldsymbol{\mu}_X| \geq \delta] = P[|Z-\mu_Z| \geq \delta] \leq \frac{\sigma_X^2}{n\delta^2}$$

➔For large *n*, $P[|Z-\boldsymbol{\mu}_X| \geq \delta] \rightarrow 0$

The pdf $f_Z(z)$ concentrates close to $\mu_Z = \boldsymbol{\mu}_X$

The End

Next time: Continue with Chp. 4