Prediction If one is interested in a random variable Y but can only observe X one wants to predict Y by a function of X, $g(X)$.

Predictor $g(X)$ is called a predictor.

To measure how good a predictor is the *mean square error* is used.

Mean square error The mean square error (mse) of g is

$$\text{mse}(g) = E[(Y - g(X))^2] \tag{1}$$

Best predictor The best predictor (with respect to minimizing the mse) is the conditional expectation $E[Y|X]$.

Conditional variance If $X = x$ is observed the conditional variance of Y is defined as

$$\begin{aligned} \text{Var}[Y|X = x] &= E[(Y - E[Y|X = x])^2/|X = x] \\ \text{Var}[Y|X] &= E[(Y|X)]^2 \end{aligned}$$

If x and Y are independent the conditional variance is equal to the regular variance.

Variance from conditional variance Calculate variance of random variable Y by observing X.

$$\text{Var}[Y] = \text{Var}[E[(Y|X)]] + E[\text{Var}[(Y|X)]] \tag{2}$$

Covariance The covariance of X and Y is given by

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])]\text{Cov}[X, Y] \quad = E[XY] - E[X]E[Y]$$

Independence and covariance X and Y are independent $\Rightarrow$ $\text{Cov}[X, Y] = 0$.
Notice that $\Leftarrow$ does NOT apply.

Increase and decrease of X and Y

- If $\text{Cov}[X, Y] > 0$ then Y increases if X is increased.
- If $\text{Cov}[X, Y] < 0$ then Y is decreased when X is increased.

Properties of the covariance

- $\text{Cov}[X, X] = \text{Var}[X]$

- Bilinearity

$$\operatorname{Cov}[aX, bY] = ab \operatorname{Cov}[X, Y]$$
$$\operatorname{Cov}[X+Y, Z] = \operatorname{Cov}[X, Z] + \operatorname{Cov}[Y, Z]$$

Variance of in- and dependent random variables 1 for independent and 2 for dependent

$$\begin{aligned} &1) \quad \operatorname{Var}[X, Y] = \operatorname{Var}[X] + \operatorname{Var}[Y] \\ &2) \quad \operatorname{Var}[X, Y] = \operatorname{Var}[X] + \operatorname{Var}[Y] + 2\operatorname{Cov}[X, Y] \end{aligned}$$

Correlation The correlation/correlation coefficient of X and Y is

$$\rho(X, Y) = \frac{\operatorname{Cov}[X, Y]}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}} \tag{3}$$

Dimensionlessness For $a, b \in \mathbb{R}$

$$\rho(aX, bY) = \frac{\operatorname{Cov}[aX, bY]}{\sqrt{\operatorname{Var}[aX]\operatorname{Var}[bY]}} = \frac{\operatorname{Cov}[X, Y]}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}} = \rho(X, Y) \tag{4}$$

Properties of the correlation coefficient

1. $-1 \leq \rho \leq 1$
2. If X and Y are independent, $\rho(X, Y) = 0$
3. $\rho = 1 \Leftrightarrow Y = aX + b$ for $b \in \mathbb{R}$ and $a > 0$
4. $\rho = -1 \Leftrightarrow Y = aX + b$ for $b \in \mathbb{R}$ and $a < 0$

The best linear predictor Let X and Y be random variables with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and correlation coefficient ρ. The best linear predictor of Y based on X is

$$l(X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_Y) \tag{5}$$

Coefficient of determination Measures how much of the variation in Y can be explained by a linear relationship of X.

$$\rho^2 = \frac{\operatorname{Var}[l(X)]}{\operatorname{Var}[Y]} \tag{6}$$

The bivariate normal distribution If (X, Y) has joint pdf

$$\begin{aligned} f(x, y) = &\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\cdot \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right] \end{aligned}$$

for $x, y \in \mathbb{R}$, then (X, Y) is said to have a bivariate normal distribution.

Properties If (X, Y) is a two-dimensional normal distribution

- $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$; $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$
- $\rho = \rho(X, Y)$
- $Y/X = x \sim \mathcal{N}\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_Y), \sigma_Y^2(1 - \rho^2)\right)$

Let (X, Y) be a bivariate normal distribution and let $a, b \in \mathbb{R}$. Then

$$aX + bY \sim \mathcal{N}\left(a\mu_X + b\mu_Y, a^2\sigma_X + b^2\sigma_Y^2 + 2ab\,\mathrm{Cov}[X, Y]\right) \tag{7}$$

If X and Y are independent and $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, then for $a, b \in \mathbb{R}$

$$aY + bY \sim \mathcal{N}\left(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2\right) \tag{8}$$