Statistical basics – A short overview (discrete)

The most important terms and definitions:

1) Expectation
2) Variance / Standard deviation
3) Sample variance
4) Covariance
5) Correlation coefficient
6) Independence vs. uncorrelation
7) Normal distribution and standard normal distribution

For comments and calculations see the appendix.

To 1)

Generally: Let g(X) be a unique function of the random variable X, then g(X) is a random variable, too. In the discrete case we define the expectation of g(X) as follows:

$$\mathbf{E[g(X)]} = \sum_{k=1}^{n} \mathbf{p_k} \cdot \mathbf{g(x_k)}\ .$$

The expectation of the random variable X itself is obtained by setting g(X)=X:

$$\mu_X = \mathbf{E[X]} = \sum_{k=1}^{n} \mathbf{p_k} \cdot \mathbf{x_k}\ .$$

Example, see appendix: [i]

In the lecture we considered the expectation of a portfolio with proportions x_i invested in stock i. Further the return of stock i is denoted by r_i.

The expected portfolio return is given by:

$$\mathbf{E[X]} = \mu = \sum_{i=1}^{n} \mathbf{x_i} \cdot \mathbf{r_i}\ .$$

Notice: Therefore our portfolio return is already an expectation.

Expectation for n variables (Assets), i.e. $\mathbf{E[X]} = \mu = \mathbf{x}_1 \cdot \mathbf{r}_1 + \mathbf{x}_2 \cdot \mathbf{r}_2 + \ldots + \mathbf{x}_\mathbf{n} \cdot \mathbf{r}_\mathbf{n}$.

For instance

n=1: $\mathbf{E[aX+b] = a \cdot E[X] + b}$, example, see appendix: [ii]

and n=2: $\mathbf{E[aX+bY] = a \cdot E[X] + b \cdot E[Y]}$, example, see appendix: [iii]

To 2)

The variance is defined as the average quadratic deviation

$\mathbf{V[X]} = \sigma^2 = \mathbf{E}[\mathbf{X} - \mathbf{E(X)}]^2$,

According to the „Theorem of Steiner“ the variance can also be written as:

$\mathbf{V[X]} = \sigma^2 = \mathbf{E}[\mathbf{X}^2] - [\mathbf{E(X)}]^2$.

The standard deviation is defined as the positive quadratic root of the variance:

$\sigma = \sqrt{\sigma^2}$.

To 3)

Having the arithmetic mean of the distribution $\bar{\mathbf{x}} = \frac{1}{\mathbf{n}} \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{x_i}$, the sample variance with n-1 degrees of freedom is the following:

$$\mathbf{S}^2 = \hat{\sigma}^2 = \frac{1}{\mathbf{n}-1} \sum_{\mathbf{i}=1}^{\mathbf{n}} (\mathbf{x_i} - \bar{\mathbf{x}})^2$$

It holds $\mathbf{E[S^2]} = \sigma^2$, therefore S^2 is called an unbiased estimator of the variance σ^2.

Proof, see appendix: [iv]

To 4)

The covariance measures the linear co-movement of X and Y:

$\mathbf{Cov(X,Y)} = \sigma_{\mathbf{X,Y}} = \mathbf{E}[(\mathbf{X} - \mathbf{E(X)}) \cdot (\mathbf{Y} - \mathbf{E(Y)})] = \mathbf{E[XY] - E[X] \cdot E[Y]}$

To 5)

The correlation coefficient is defined on [-1,1] and has the following form:

$$\mathbf{Corr(X,Y)} = \rho_{\mathbf{X,Y}} = \frac{\mathbf{Cov(X,Y)}}{\sqrt{\mathbf{V(X)\cdot V(Y)}}} = \frac{\sigma_{\mathbf{X,Y}}}{\sigma_{\mathbf{X}} \cdot \sigma_{\mathbf{Y}}}.$$

With 5) we receive this equation (used in the tutorial) :

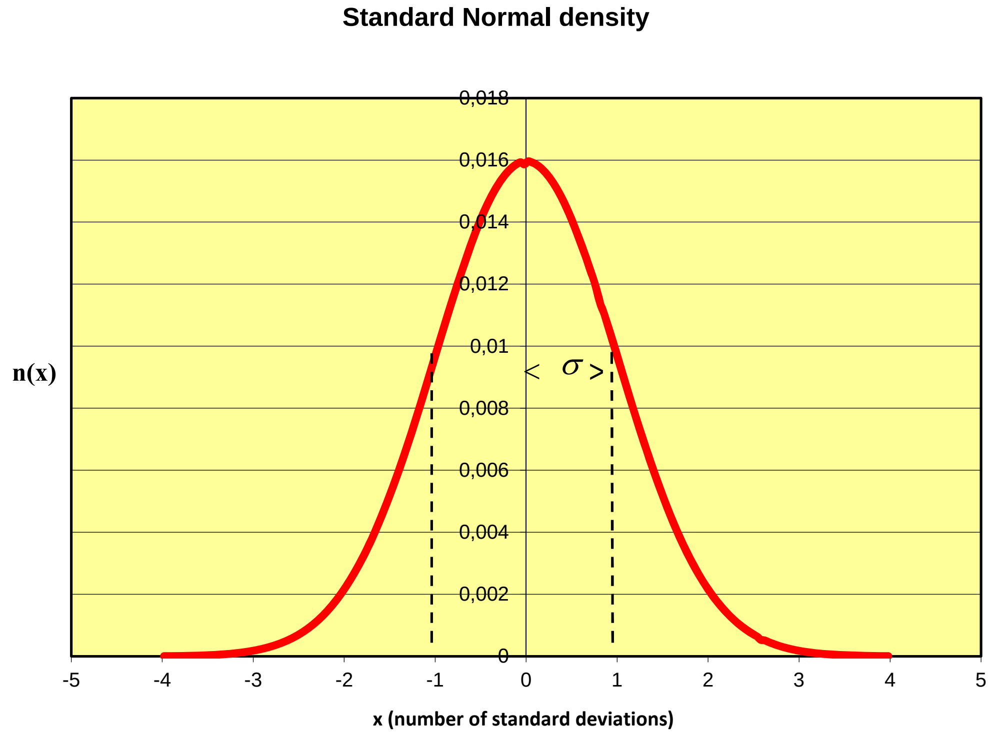
$$\mathbf{Cov(X,Y)} = \boldsymbol{\sigma}_{\mathbf{X,Y}} = \boldsymbol{\rho}_{\mathbf{X,Y}} \cdot \sqrt{\mathbf{V(X)\cdot V(Y)}} \Leftrightarrow \mathbf{Cov(X,Y)} = \rho_{\mathbf{X,Y}} \cdot \sigma_{\mathbf{X}} \cdot \sigma_{\mathbf{Y}}$$

To 6)

Generally, you cannot take the following implication: an uncorrelated random variable is also independent. This holds only for symmetric distributions, like the normal distribution.

To 7)

A random variable X is called normal distributed, for short: X ~ N (μ , σ^2), if it has a normal density function with parameters μ and σ^2.



A normal distribution with $\mu = 0$ and $\sigma^2 = 1$ is called a standard normal distribution N(0,1).

The distribution has the density function $\mathbf{n(x)} = \frac{1}{\sqrt{2\pi}}\mathbf{e}^{-\frac{x^2}{2}}$, as plotted in the graph above:

Some properties of the normal distribution:

· unimodal distribution

· symmetric distribution with maximum at x = μ

· points of inflexion at x = $\mu \pm \sigma$

· E (X) = μ, Var (X) = σ^2

You can transform any normal distribution into a standard normal distribution.

For X ~ N (μ, σ^2)-distributed random variable, $\mathbf{U} := \frac{\mathbf{x} - \mu}{\sigma}$ is a standard normal distributed random variable.

For that reason all calculations (probabilities, quantiles, a.s.o.) can be done based on a standard normal distribution and you do not need to do calculation for each N (μ, σ^2)-distributed random variable.

Let $\mathbf{x_p}$ be the quantile of order p of a N (μ, σ^2)-distribution and $\lambda_\mathbf{p}$ the quantile of order p of a N(0, 1)-distribution. Then:

$$\mathbf{x_p} = \mu + \lambda_\mathbf{p} \cdot \sigma, \forall \mathbf{p} \in (0,1).$$

By considering a normal distribution the central coverage interval lies symmetrically around the mean. Having these N(0,1) quantiles you can determine the coverage interval for a probability $1-\alpha$ for a N(μ , σ^2)-distributed random variable X as

$$\mathbf{P}(\mu - \lambda_{1-\frac{\alpha}{2}} \cdot \sigma \leq \mathbf{X} \leq \mu + \lambda_{1-\frac{\alpha}{2}} \cdot \sigma) = 1 - \alpha \text{, with } \lambda_\mathbf{p} = -\lambda_{1-\mathbf{p}}.$$

The other way round you can calculate out of the quantiles, the corresponding probability of the realization lying between the $\mu \pm \lambda_{1-\frac{\alpha}{2}}\sigma$ interval. That probability can be interpreted as the relative frequency.

For $\lambda_{1-\frac{\alpha}{2}}$ = K we receive the relative frequency for the K* σ -bands. For instance:

K=1: $P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6827$

i.e. approximately 68% of all normal realisations lie within the band $\mu \pm \sigma$.

K=2: $\mathbf{P}(\mu - 2\sigma \leq \mathbf{X} \leq \mu + 2\sigma) = 0{,}9545$

i.e. approximately 95% of all normal realisations lie within the band $\mu \pm 2\sigma$.

K=3: $$\mathbf{P}(\mu - 3\sigma \leq \mathbf{X} \leq \mu + 3\sigma) = 0{,}9973$$

i.e. approximately 99.7% of all normal realisations lie within the band $\mu \pm 3\sigma$.

Appendix:

---

i *E.g.* expectation of a discrete distribution function P (X = x_i):

$$\mathbf{E[X]} = \sum_{i=1}^{n} x_i \cdot p(x_i) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

ii *example* (chapter risk and return, slide 8):

Assumption: A discrete random variable X with density function f(x), and constants a and b.

Proposition: $\mathbf{E[aX + b] = a \cdot E[X] + b}$

Prove: $$\mathbf{E[aX+b]} = \sum_{i=1}^{m} (ax_i + b) \cdot f(x_i)$$

$$= \sum_{i} ax_i f(x_i) + bf(x_i)$$

$$= a\sum_{i} x_i f(x_i) + b\sum_{i} f(x_i)$$

$$= aE[X] + b.$$

q.e.d.

You can show $\mathbf{V[aX + b] = a^2 \cdot V[X]}$ in the same way.

iii As well as for the variance: $\mathbf{V[aX + bY] = a^2 \cdot V[X] + b^2 V[Y] + 2 \cdot a \cdot b \cdot Cov(X, Y)}$

$$\Leftrightarrow \mathbf{V[aX + bY] = a^2 \cdot V[X] + b^2 V[Y]} + 2 \cdot \mathbf{a} \cdot \mathbf{b} \cdot \sigma_X \cdot \sigma_Y \cdot \rho_{X;Y}$$

iv *calculation :*

$$\mathbf{E[S^2]} = \mathbf{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

$$= \frac{n}{n-1} \cdot \frac{1}{n} \mathbf{E}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

$$= \frac{n}{n-1} \cdot \frac{1}{n}\sum_{i} \mathbf{E}[(x_i - \bar{x})^2] = \frac{n}{n-1} \cdot \frac{1}{n}\sum_{i} \mathbf{E}[(\{x_i - \mu\} - \{\bar{x} - \mu\})^2]$$

with the binomial formula and $\mathbf{E}[(\bar{x} - \mu)] = \frac{\sigma}{n}$ it follows:

$$\mathbf{E[S^2]} = \frac{n}{n-1} \cdot \frac{1}{n}\left[n \cdot \sigma^2 + \frac{n\sigma^2}{n} - 2 \cdot \mathbf{E}[(\bar{x} - \mu)\sum (x_i - \mu)]\right]$$

$$= \frac{n}{n-1}\left[\sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n}\right] = \frac{n}{n-1}\left[1 + \frac{1}{n} - \frac{2}{n}\right]\sigma^2 = \sigma^2.$$

q.e.d.