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) Independency vs. No correlation
7) Normal distribution and standard normal distribution

Comments and calculations see the appendix.

To 1)

Generally:

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

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

Getting the expectation of the random variable X itself, we set g(X)=X and receive:

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

Example, see appendix: [i]

Lecture: Considering expectations of a portfolio with weight x_i in stock i invested and the returns r_i:

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

Notice: therefore our portfolio returns are already expectations.

Expectation is given for n variables, i.e. $E[X] = \mu = x_1 \cdot r_1 + x_2 \cdot r_2 + ... + x_n \cdot r_n$.

For instance

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

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

To 2)

The variance is defined as the average quadratic deviation

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

resp.: „Theorem of Steiner" :

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

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

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

$E[S^2] = \sigma^2$ always holds true,

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:

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

To 5)

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

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

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

$Cov(X,Y) = \sigma_{X,Y} = \rho_{X,Y} \cdot V(X) \cdot V(Y) \Leftrightarrow Cov(X,Y) = \rho_{X,Y} \cdot \sigma_X \cdot \sigma_Y$

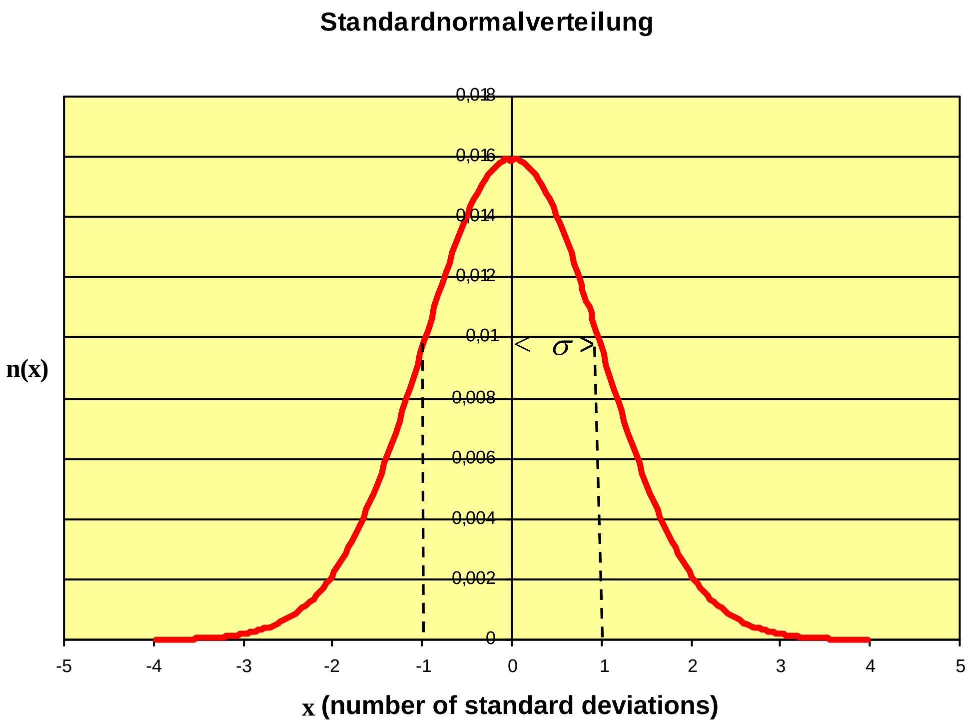
To 6)

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

Zu 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) having the following density function $n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, as shown above:



Some properties of the normal distribution:

· unimodal distribution

· symmetric distribution with maximum at x = μ

· point 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, $U = \frac{x - \mu}{\sigma}$ is a standard normally distributed one.

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

Let x_p be the quantile of order p of a N (μ , σ^2)- distributed random variable and λ_p the quantile of order p of a N(0, 1) - distributed random variable. Then:

$$x_p = \mu + \lambda_p \cdot \sigma, \forall p \in (0,1).$$

The central coverage interval lies symmetrically around the mean assuming a symmetric distribution.

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

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

Having the probability $1-\alpha$ you can determine the quantiles for the corresponding interval; the other way round you can calculate out of the quantiles the corresponding probability, which may be interpreted as the relative frequency.

Consider $\lambda_{1-\frac{\alpha}{2}}$ for 1, 2, 3, ... , we receive the above kσ -bands:

K=1: $P(\mu - \sigma \le X \le \mu + \sigma) = 0{,}6827$

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

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

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

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

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):

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

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

It is given: A discrete random variable X with density function f(x), and two constants a and b.

Show: $E[aX + b] = a \cdot E[X] + b$

Solution:
$$\begin{aligned} E[aX + b] &= \sum_{i=1}^{m} (ax_i + b) \cdot f(x_i) \\ &= \sum_i ax_i f(x_i) + bf(x_i) \\ &= a\sum_i x_i f(x_i) + b\sum_i f(x_i) \\ &= aE[X] + b. \end{aligned}$$

q.e.d.

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

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

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

iv *calculation :*

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

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

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

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

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

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

q.e.d.