Statistische Grundlagen – Ein kurzer Überblick (diskret)

Die wichtigsten Begriffe und Definitionen:

1) Erwartungswert
2) Varianz / Standardabweichung
3) Stichprobenvarianz
4) Kovarianz
5) Korrelationskoeffizient
6) Unabhängigkeit vs. Unkorreliertheit
7) Normalverteilung und Standardnormalverteilung

Erklärungen und Rechnungen im Anhang.

Zu 1)

Ganz allgemein: Wenn g(X) eine eindeutige Funktion der Zufallsvariablen X ist, dann ist auch g(X) eine Zufallsvariable. Der Erwartungswert der Zufallsvariablen g(X) wird im diskreten Fall definiert durch:

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

Den Erwartungswert der Zufallsgröße X selbst erhält man mit g(X)=X:

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

Beispiel siehe Anhang: [i]

In der Vorlesung betrachten wir die den Erwartungswert eines Portfolios mit Anteilen x_i investiert in Aktie i. Weiter bezeichnet r_i die Rendite der Aktie i

Der erwartete Portfolioreturn ist gegeben durch:

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

Bem.: Die Rendite unseres Portfolio ist somit schon ein Erwartungswert.

Der Erwartungswert für n Variablen (Aktien) , d.h. $\mathbf{E[X]} = \mu = \mathbf{x}_1 \cdot \mathbf{r}_1 + \mathbf{x}_2 \cdot \mathbf{r}_2 + \mathbf{...} + \mathbf{x_n} \cdot \mathbf{r_n}$.

Betrachte im folgenden den Fall für

n=1: $\mathbf{E[aX + b] = a \cdot E[X] + b}$, Beispiel dazu siehe Anhang: [ii]

und n=2: $\mathbf{E[aX + bY] = a \cdot E[X] + bE[Y]}$, Beispiel dazu siehe Anhang: [iii]

Zu 2)

Die Varianz ist die mittlere quadratische Abweichung

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

Nach dem „Satz von Steiner“ kann die Varianz auch geschrieben werden als:

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

Die Standardabweichung ist definiert als die positive Quadratwurzel der Varianz

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

Zu 3)

Mit dem arithmetischen Mittel der Verteilung $\bar{\mathbf{x}} = \frac{1}{\mathbf{n}} \sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{x_i}$ lautet die Stichprobenvarianz mit n-1 Freiheitsgraden:

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

Es gilt: $\mathbf{E[S^2]} = \sigma^2$, damit ist S^2 ein unverzerrter Schätzer für die Varianz σ^2.
Beweis siehe Anhang: [iv]

Zu 4)

Die Kovarianz ist ein Maß für den linearen Zusammenhang von X und Y und lautet:

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

Zu 5)

Der Korrelationskoeffizient ist auf das Intervall [-1,1] normiert und hat die folgende Form

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

Aus 5) ergibt sich die Umformung, die in den Übungen verwendet wird:

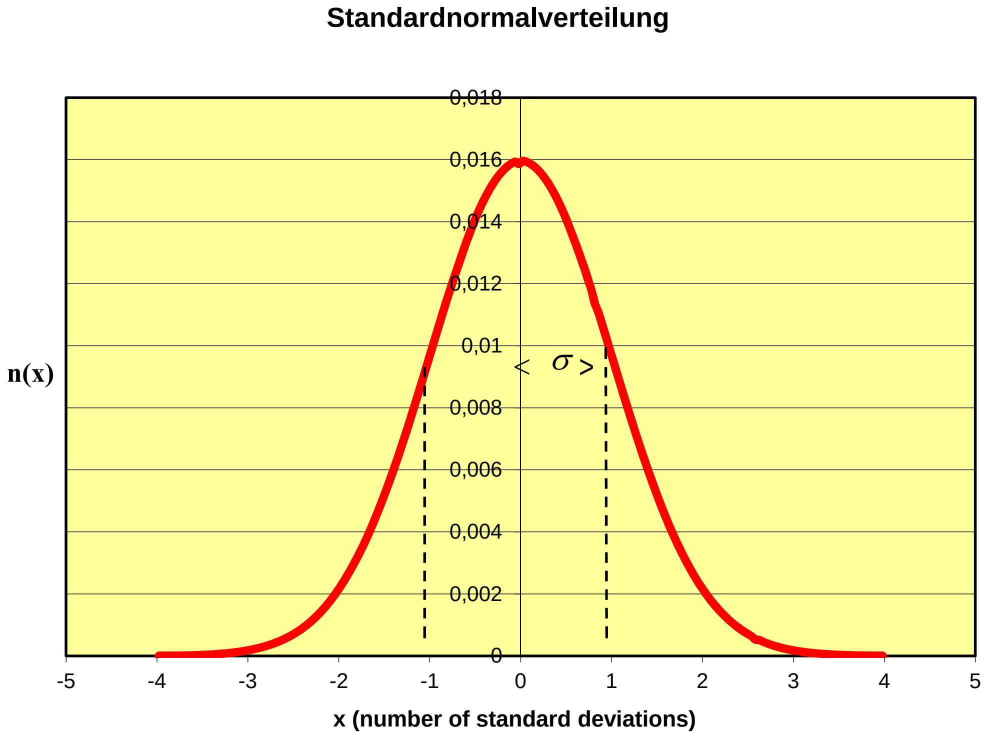
$$\mathbf{Cov(X,Y)} = \boldsymbol{\sigma}_{\mathbf{X,Y}} = \boldsymbol{\rho}_{\mathbf{X,Y}} \cdot \sqrt{\mathbf{V(X)\cdot V(Y)}} \Leftrightarrow \mathbf{Cov(X,Y)} = \rho_{\mathbf{X,Y}} \cdot \sigma_{\mathbf{X}} \cdot \sigma_{\mathbf{Y}}$$

Zu 6)

Aus der Unkorreliertheit folgt im Allgemeinen noch nicht, dass die Zufallsvariablen auch unabhängig sind. Dies ist nur bei symmetrischen Verteilungen wie z.B. der Normalverteilung der Fall.

Zu 7)

Eine Zufallsvariable X heißt normalverteilt, kurz X ~ N (μ , σ^2), falls ihre Dichte normalverteilt ist mit Parametern μ und σ^2.



Eine Normalverteilt mit $\mu = 0$ und $\sigma^2 = 1$ heißt Standardnormalverteilung, kurz N (0,1).

Sie besitzt die Dichtefunktion $\mathbf{n(x)} = \frac{1}{\sqrt{2\pi}}\mathbf{e}^{-\frac{x^2}{2}}$, wie in dem Graph oben gezeichnet.

Einige Eigenschaften der Normalverteilung:

· unimodale Verteilung

· symmetrische Verteilung mit Maximum bei x = μ

· Wendepunkte bei x = $\mu \pm \sigma$

· E (X) = μ , Var (X) = σ^2

Jede Normalverteilung kann in eine Standardnormalverteilung transformiert werden.

Sei X eine Normalverteilte Zufallsvariable, X ~ N (μ , σ^2). Dann ist die Transformation $\mathbf{U} := \frac{\mathbf{x} - \mu}{\sigma}$,

eine standardnormalverteilte Zufallsvariable. Aus diesem Grund können alle Berechnungen (Wahrscheinlichkeiten, quantile, usw.) basierend auf einer Standardnormalverteiltung durchgeführt werden und müssen nicht für jede N (μ , σ^2)-Verteilte Zufallsvariable einzeln berechnet werden.

Sei $\mathbf{x_p}$ das Quantil einer Normalverteilung mit Parametern μ und σ^2 von der Ordnung p und sei $\lambda_\mathbf{p}$ das Quantil einer Standardnormalverteilung (μ =0 und σ^2 =1). Dann gilt:

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

Bei einer Normalverteilung liegt das Konfidenzintervall symmetrisch um den Erwartungswert der Verteilung. Das Konfidenzintervall, einer N(μ , σ^2) verteilten Zufallsvariable, zum Signifikanzniveau α kann aus den Quantilen der N(0,1)-Verteilung bestimmt werden. Es gilt

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

Andererseits lässt sich für gegebene Quantile die Wahrscheinlichkeit bestimmen, dass die Realisationen der Verteilung innerhalb des Intervalls $\mu \pm \lambda_{1-\frac{\alpha}{2}}\sigma$ liegen. Diese Wahrscheinlichkeit kann als relative Häufigkeit interpretiert werden.

Für $\lambda_{1-\frac{\alpha}{2}}$ = K erhalten wir die relative Häufigkeit bei einem Interquartilsabstand von K* σ . Zum Beispiel:

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

d.h. ungefähr 68% aller Realisationen der Normalverteilung liegen innerhalb des Intervalls $\mu \pm \sigma$.

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

d.h. ungefähr 95% aller Realisationen der Normalverteilung liegen innerhalb des Intervalls $\mu \pm 2\sigma$.

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

d.h. ungefähr 99.7% aller Realisationen der Normalverteilung liegen innerhalb des Intervalls $\mu \pm 3\sigma$.

Anhang:

---

i Ein *Beispiel* dafür ist der Erwartungswert einer Verteilungsfunktion P (X = x_i) im diskreten Fall:

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

ii *Beispiel zu Folie 8* aus Risiko und Ertrag:

Vorraussetzung: Diskrete Zufallsvariable X mit Dichte f(x), und die Konstanten a und b.

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

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

$$\mathbf{=\sum_i ax_i f(x_i)+bf(x_i)}$$

$$\mathbf{=a\sum_i x_i f(x_i)+b\sum_i f(x_i)}$$

$$\mathbf{=aE[X]+b.}$$

q.e.d.

Analog lässt sich der Zusammenhang für die Varianz $\mathbf{V[aX+b]=a^2\cdot V[X]}$ zeigen.

iii Ebenso für die Varianz: $\mathbf{V[aX+bY]=a^2\cdot V[X]+b^2V[Y]+2\cdot a\cdot b\cdot Cov(X,Y)}$

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

iv *Rechnung dazu:*

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

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

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

mit Binomischer Formel und $\mathbf{E[(\bar{x}-\mu)]=\frac{\sigma}{n}}$ folgt

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

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

q.e.d.