Grundlagen der Informatik

Me

28. Januar 2021

Inhaltsverzeichnis

1	Lineare Optimierung					
	1.1	Simplexalgorihmus	4			
2	Finanzmathematik					
	2.1	Folgen und Reihen	5			
	2.2	Einfache Verzinsung	5			
		2.2.1 Verzinsung mit Zinseszins	5			
		2.2.2 Stetige Verzinsung	5			
		2.2.3 Gemischte Verzinsung	6			
		2.2.4 Freigeld	6			
	2.3	Investitions- und Finanzrechnung	6			
	2.4	Rentenrechnung	6			
		2.4.1 Zinsperiode = Ratenperiode	6			
		2.4.2 Rentenperiode < Zinsperiode	7			
	2.5	Tilgungsrechnung	7			
		2.5.1 jährliche Ratentilgung	7			
		2.5.2 unterjährliche Ratentilgung	7			
3	Des	kriptive Statistik	7			
	3.1	Stichproben	7			
	3.2	Häufigkeiten, empirische Verteilungsfunktion	8			
	3.3	Klasseneinteilung	8			
	3.4	Statistische Parameter	9			
	3.5	Zweidimensionale Stichproben	9			
	3.6		10			
4	Kon	nbinatorik 1	L 1			
•	4.1		11			
	4.2	9	11			
	4.3		11			
	4.4		11			
5	Wał	hrscheinlichkeitstheorie 1	12			
J	5.1					
	5.2	· · · · · · · · · · · · · · · · · · ·	12			
	5.3	•	13			
	5.4	<u> </u>	13			
	5.5		13			
	5.6		13			
6	7uf:	allsvariable und statistische Verteilung	13			
	6.1	——————————————————————————————————————	13			
	6.2		13			

7	Indu	ktive Statistik und statistische Tests	13
	6.11	Exponentially erteilung	13
		Normalverteilung	
		stetige Gleichverteilung	
	6.8	stetige Zufallsvariable	13
	6.7	Poisson-Verteilung	13
	6.6	Hypergeometrische Verteilung	13
	6.5	Geometrische Verteilung	13
	6.4	Binomialverteilung	13
	6.3	Diskrete Gleichverteilung und Bernoulli-Verteilung	13

1 Lineare Optimierung

Unter Linerare Optimierung versteht man, dass eine bestimmte Funktion durch einstellen bestimmter 'Parameter' möglichst groß (oft für Gewinn benutzt) oder möglichst klein (für Verluste) wird. Dafür wird die öineare Zielfunktion:

$$g(\vec{x}) = g_0 + g_1 x_1 \dots g_n x_n$$

hergenommen. Zum Beispiel ist g_i dann der Preis für die Herstellung eines Produkts und x_i dann die Anzahl wie oft das Produkt hersgetellt werden soll und x_i sind eben die Parameter die eingestellt werden sollen. Dazu sind idR. Randbedingungen gegeben, wie ' x_1 darf höchstens doppelt so oft hergestellt werden wie x_2 '. Dann gibt es noch triviale Randbedingungen. Da man nicht negativ produzieren kann gilt meistens auch $x_i \geq 0$. Somit erhält man einige Ungleichungen und Gleichung. Zuerst schriebt man alle (Un)Hleichungen untereinander auf und drhet jedes Ungleichungszeichen (Multiplikation mit -1) um damit alle Ungleichungszeichen in die selbe Richtung zeigen. Nun versucht man Variablen zu eliminieren indem man de GLeichungen die man hat nach einer Variable umstellt und dann diese Variable in den Ungleichungen ersetzt. Beispiel:

$$x_1 + x_2 = 0 \Rightarrow x_1 = -x_2$$

 x_1 kan in diesem Beispiel mit $-x_2$ in den Ungleichungen ersetzt werden. Wenn man die ganzen Gleichungen in ein Diagramm einzeichnet erhält man ein Vieleck mit Kanten die eben von den Ungleichungen darsgetellt werden. Man kann jetzt jeden Schnittpunkt anschauen, die Werte berechnen die herauskommen wenn man die x-Koordianten in die Gewinngleichung einsetzt und vergleicht wie jeden Wert und sucht sich den höchsten/nidrigsten Wert als Lösung heraus. Bei vielen Parametern kann das schnell viele Berechnungen benötigen. Man kann auch das Diagramm wieder hernehmen. Hier sucht man sich einen geigneten Startpunkt (eine Punkt an einer Kante) und schaut ob man ob die Kante ansteigt oder waagrecht verläuft und folgt dieser Kante bis man einen Schnittpunkt findet. Hier wird wiederum geschaut ob es eine Kante héibt die ansteigt oder waagrecht verläuft. Dies wird wiederholt bis das nicht mehr möglich ist. Der Letzt gefundene Schnittpunkt, welcher noch das Vieleck umspannt, erhält dann die Parameter für die Lösung.

1.1 Simplexalgorihmus

Da man schwer vier, fünf oder sechs Dimensionen zeichnen kann gibt es den Simplexalgorihmus welcher den Optimalwert rechnerisch ermitteln kann. Hier muss zuerst ein Tabloet

erstellt werden. Die Form des Tablots ist wie folgt:
$$\begin{pmatrix} A & \vec{k} \\ \hline \vec{g} & G \end{pmatrix}$$
 Wobei Die Matrix A Die Ungleichungen ind Marixform ist, $\vec{g} = G$ eine Lösung ist von der man weiß, dass sie

Die Ungleichungen ind Marixform ist, $\vec{g} = G$ eine Lösung ist von der man weiß, dass sie die Gewinnfunktion erfüllt und sich $\vec{k} = A\vec{u} - \vec{b}$. Dabei ist \vec{u} auch der Vektor der die Gewinnfunktion erfüllt.

2 Finanzmathematik

2.1 Folgen und Reihen

Eine Zahlenfolge ist eine Abbildung, welche wie folgt defineiert ist:

$$a = \mathbb{N} \to \mathbb{R}$$

Anstatt a(i) schriebt man a_i für jedes einzelne Glied der Folge. Eine Reihe ist eine Aufsummierung aller Folgeglieder und wird unendliche Reihe genannt wenn unendlich viele Glieder aufsummiert werden bzw. endliche Reihe genannt wenn bis zu einem bestimmten index aufsummiert wird.

arithmetische Zahlenfolge: $a_n = a_1 + (n-1)d, d\epsilon \mathbb{R}$ fest

endliche arithmetische Reihe: $\frac{N(2a_1 + (N+1)d)}{2}$

geometrische Zahlenfolge: $a_n = a_1 \cdot q^{n-1}$

geometrische Reihe: $a_1 \frac{q^N - 1}{q - 1}$

2.2 Einfache Verzinsung

Sollen Zinsen nicht miteinander Verzinst werden (Zinsen werden auf ein anderes Konto ausgezahlt) spircht ma nvon einfacher Verzinsung. Es gilt folgenden Formel:

$$K_n = K_0(1 + n \cdot i)$$

Wobei K_n das Kapitale nach n Zinsperioden dasrstellt. K_0 ist dabei das Startkapital, i die Zinsrate¹ und n eben die Laufzeit in Zinsperioden (meistens Monate oder Jahre).

2.2.1 Verzinsung mit Zinseszins

Hier werden die Zinsen mitverzinst (Zinsen werden aufs selbe Konto ausgezahlt). Die Formel lautet:

$$K_n = K_0(1+i)^n$$

2.2.2 Stetige Verzinsung

Unter stetiger Verzinsung versteht man, dass die Zinsen nicht jeden Monat oder jeden Tag angerechnet werden sondern zu jedem Augenblick. Dafür wird folgende Fromel werwendet:

$$K_n = K_0 e^{i \cdot n}$$

 $^{^{1}}i \cdot 100 \equiv ZinsrateinProzent$

2.2.3 Gemischte Verzinsung

Wenn man als Zisnperiode ein gnazes Jahr hat aber man die Verzinsung mit en in der Zinsperiode 'abbrechen' will wird folgende Formel benutzt:

$$K_n = K_0(1+i)^{\lfloor n \rfloor} \cdot (1+i(n-\lfloor n \rfloor)^2)$$

2.2.4 Freigeld

Die Idee hinter Freigeld ist dass Geld nicht wir Lebensmittel oä. verderben kann. Um das zu korrigieren kann man die Formel für einfache Verzinsung (wenn man Bargeld hat) bzw. die Formel für den Zinseszins hernehmen (Girokonto). Hier nimmt man als Zinsrate einfach den Negativwert.

2.3 Investitions- und Finanzrechnung

Die Investionsrechnung wird verwendet um zu bestimmen ob sich eine Investions im Laufe der Zeit lohnen wird. Zuerst zieht man die Kosten der Investions vom Kapital ab. Gegenfalls schaut man ob es periodische Zahlungen gibt die man leisten muss (Wartungsarbeiten, Kreditszinsen etc.) und zieht diese auch ab. Man schätz nun den Gewinn den die Investion im Monat/Jahr etc. macht, zieht noch eventuelle Zahlungen ab und erhält den Monats/Jahresgewinn. Dies macht man für alle Monate für die eine Schätzung machbar ist, verzinst diese dementsprechend und vergleicht den Gesamtgewinn mit dem Gewinn den man erhält wenn man nicht investiert mit dem über dem selben verzinsten Kapital, das man hat wenn man nicht investiert.

2.4 Rentenrechnung

Unter einer Renter versteht man im Allgemeinem eine regelmäßige Ein- oder Auszahlung. Es wird zwischen auch noch eine Unterschiedung getroffen wenn die Zinsperiode der Rentenperiode gleicht (man zahl jeden Monat einen gewissen Betrag aufs Sparbuch und bekommt jeden Monat auch Zinsen auf das Geld) und wenn die Rentenperiode kleiner der Zinsperiode ist (man zahlt monatlich ein bekommt aber nur jährlich Zinsen).

2.4.1 Zinsperiode = Ratenperiode

Die allgemeine Formel lautet:

$$R_n = r \cdot q \cdot \frac{q^n - 1}{q - 1}$$

r ist die Ratenzahlung, n wieder die Anzhal der Ratenperioden, q berechnet sich durch q=1+i, wobei i wieder der Zins ist. Die Formel gilt für vorschüssige Ratenzahlungen und nachschüssige Zinszahlungen. Für die nachschüssige Ratenzahlung gilt:

$$R_n = r \cdot \frac{q^n - 1}{q - 1}$$

²Die Klammern sind Gaussche Rundungsklammern. Sind sie wie hier nach unten gerichtet, wird abgerundet und wenn sie nach oben gerichtet sind, wird aufgerundet

2.4.2 Rentenperiode < Zinsperiode

Wenn die Rentenperiode kleiner der Zinsperiode ist muss man nur die Ratenzahlung r aus der Formel für die nachschüssige Ratenzahlung ändern. Es gilt für das neue r_e :

$$r_e = r \cdot (m + \frac{i}{2} \cdot (m+1)) \Rightarrow R_n = r_e \frac{q^n - 1}{q - 1} = r \cdot (m + \frac{i}{2} \cdot (m+1)) \cdot \frac{q^n - 1}{q - 1}$$

Wobei das m die Anzahl der Ratenzahlungen pro Zinsperiode ist.

2.5 Tilgungsrechnung

Bei der Tilgunsrechnung geht es idR. umd die Rückzahlung eines Krdited (bzw. allgemein Schulden). Die Annuität bezeichnet damit den Betrag, der periodisch zurückgezahlt werden muss. Die Annuität setzt sich wiederum aus der Tilgungsrate (Betrag der zurückgezahlt werden soll) und dem Zinsbestandteil (Zinsen die pro Periode von den Restschulden dazukommen).

2.5.1 jährliche Ratentilgung

Soll die Tilgunsrate jedes Jahr (Periode) gleich bleiben ergibt sich folgende Formel:

$$T = \frac{S}{n}$$

Wobei T die Gesamtschuld ist und n die Anzahl der Jahre (perioden) der Kreditlaufzeit. Will man die Zinsrate für ein bestimmtes Jahr (hier r) ausrechnen gilt folgende Formel:

$$Z_r = T \cdot (n - r + 1) \cdot i$$

i ist dabei wieder der Zinsatz,
n Die Kreditlaufzeit und T die Tilgungsrate.

Die Annuität ergibt sich dann wenn man die beiden speziellen Formeln nun in die allgemeine Formel von oben einsetzt und umformt:

$$A = T \cdot (1 + (n - r + 1)i)$$

2.5.2 unterjährliche Ratentilgung

Soll mehrmals im Jahr getilgt werden muss der jährliche Zinssatz durch die Anzahl der Tilgungen im Jahr geteilt werden. Weiter beschreibt das n nicht mehr das ahr sondern die Anzahl der Zinsperioden.

3 Deskriptive Statistik

3.1 Stichproben

Bei der Statistik versucht man allgemeine Aussagen über einen bestimmten Datensatz zu erheben. Bei Millionen von Daten ist es oft nicht sinnvoll alle Daten einzeln zu betrachten. Die Datenmenge wird auch als Beobachtungsmenge (B) bezeichnet. Jedem Element

aus B wird dann ein bestimmtes Merkmal zugeordnet. Die Merkmale können disktret³ oder stetig⁴ sein. Weiter ist Urliste eine Teilmenge aus B mit den zugehörigen Merkmalswerten. Die Merkmalswerte aus der Urliste bilden dann die Stichprobe, welche geordnet oder ungeordnet sein kann. Der Stichprobenumganf (n) ist dann diue Kardinalität der Stichprobenmenge.

3.2 Häufigkeiten, empirische Verteilungsfunktion

Bei Häufigkeiten wird zwischen der absoluten und relativen Haüfigkeiten unterschieden. Dafür nimmt man die zuerst die Funktion g(x) her, welche zählt wie oft ein Element x in der Stichprobe vorkommt. Dies ist dann die absolute Häufigkeit. Teilt man die absolute Häufigkeit durch n^5 .

absolute Häufigkeit:
$$g(x) \neq 0 \Leftrightarrow x \in \{x_1, x_2, ..., x_n\} \subset \mathbb{R} (\mathbb{N})$$
 relative Häufigkeit: $h(x) = \frac{1}{n}g(x)$

Die Summenhäufigkeit (G(x)) summiert alle Werte aus der Stichprobe bis einschließlich x. Die relative Häufigkeit nimmt die Summenhäufigkeit und teilt diese durch den Stichprobenumfang.

Summenhäufigkeit
$$G(x_i) = x_1 + x_2 + x_3 + \dots + x_i | x_i < x_n$$
relative Summenhäufigkeit:
$$H(x) = \frac{1}{n} G(x)$$

Die realtive Summenhäufigkeit wird auch die empirische Verteilfunktion genannt. Durch die Summenhüfigkeit werden Häufigkeiten zusammengefasst und man kann allgemeinere Aussagen für die zusammengefassten Häufigkeiten machen.

3.3 Klasseneinteilung

Oft ist es sinnvoll Stichporbenwerte in Klassen einzuteilen. Dafür wählt man zuerst die Anzahl an Klassen die man haben möchte (hie: l). Nun teilt man die den Stichprobenumfang durch l. Dadurch erhält man die Klassenbreite w (ggf. muss gerundet werden). Die Klassenbreite gibt an wie viele Einzelemenet zusammengefasst werden soll (durch Addition). Hat man alle Klassen zusammengefasst kann man nun allgemeinere Aussagen über bestimmte Teile der Stichprobe machen (Aussagen über den Verdienst von Personen die weniger als 20h, zwischen 20h und 40h und mehr als 40h die Woche arbeiten als Beispiel).

Für die Klasseneinteilung sollte man sich an diese Regeln halten:

• Möglichst gleich große Klassen

 $^{^3\}mathbb{N}_0$ also alles was abzählbar ist wie Gehalt oder Anzahl von Besuchern etc.

 $^{^4}$ Intervalle in $\mathbb R,$ beleibig genau ermittelbare Werte wie Zeit Länge

⁵Stichprobenumfang

- $5 \le l^6 \le 25$
- $1 \approx \sqrt{n}$
- Klassengrenzen solten ganze Zahlen sein
- Klassen sollten den Wertebereich der Stichprobe möglichst gut abdecken

3.4 Statistische Parameter

Desto allgemeiner man eine Stichprobe behandelt, desto übersichtlicher aber auch desto ungenauer werden die Informationen. Es gibt statistische Paramter welche die Stichprobe in eine einzelne Zahl zusammenfassen:

arithmetisches Mittel/Stichprobenmitte:
$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$
 Median für n = ungerade:
$$\tilde{x} = x_{(\frac{n+1}{2})}$$
 Median für n = gerade:
$$\tilde{x} = \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$$
 Spannweite:
$$v := x_{(n)} - x_{(1)}$$
 Varianz:
$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$
 Streuung/Standardabweichung:
$$s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

3.5 Zweidimensionale Stichproben

Eine Zweidimensionale Stiichprobe ist wie eine normale Stichporbe nur dass da jetzt noch ein weiteres Merkmal auftritt, wodurch Zahlenpaare entstehen (zB. Antatt nur Einkommen als Mekrmal wird jetzt Einkommen und Steuern untersuch). Für zweidimensionale

⁶Anzahl der Klassen

Stichproben gelten (mit kleinen Änderungen) die selben Definitionen:

Stichenprobenumfang: $n = \text{Anzahl der } (\mathbf{x}, \mathbf{y}) \text{ Zahlenpaare absolute Häufigkeit:} \qquad g(x,y) \neq 0 \Leftrightarrow x,y \in \{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\}$ relative Häufigkeit: $h(x,y) = \frac{1}{n}g(x,y)$ Summenhäufigkeit: $G(x,y) = \sum_{\{(a,b): a \leq x, b \leq y\}} g(a,b)$ empirische Vertielungsunktion: $H(x,y) = \sum_{\{(a,b): a \leq x, b \leq y\}} h(a,b)$ Klassenhäufigkeiten: $\tilde{h_{ij}} = \sum_{x \in KX_i, y \in KY_j} h(x,y)$ Randhäufikeiten für ein festes x/y: $h_{x/y}(x/y) = \sum_{x/y \in \mathbb{R}} h(x,y)^7$

3.6 Lineare Regression

Zeichnet man die Stichprobenwerte in ein Diagramm ein erählt man eineige Punkt in einem Diagramm. Mit der linearen Regression sucht man eine Funtion, welche ungefähr der Richtung der Punkten entspricht. Für die Funktion y=ax+b müssen a und b so gewählt werden, damit der Abstand von der Geraden zu den Punkten mäglichst klein wird. Die einzelnen Abstände Δ werden dann quadriert und danach zusammenaddiert. a und b sollen so gewählt werden, dass Δ^2 möglichst klein wird. Diese Methode zur Ermittlung der Regressionsgeraden heißt Methode der kleinsten Quadrate". Die Kovarianz bei zweidimensionalen Stichproben wird wie folgt ermittelt:

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n-1} \sum_{i=1}^{n} x_i y_i - n \overline{x} \overline{y}$$

Teilt man die Kovarianz durch die Standardabweichung von jeweils x und y erhält man den empirischen Korrelationskoeffizienten:

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_x}$$

Weiter gilt für den Korrelationskoeffizienten r_{xy} :

- $-1 \le r_{xy} \le 1$
- Wenn $r_{xy} = 1$ oder $r_{xy} = -1$ liegen die Messpunkte direkt auf einer Geraden
- \bullet Wenn $r_{xy}=0$ dann haben die Mekrmale X und Ykeinen Zusammenhang
- Gibt es mehr als 2 Merkmale kann man selber Paare bilden und sieht dadurch wie stark Merkmale direkt miteinander zusammenhängen

4 Kombinatorik

4.1 Einführung

Bei der Kombinatorik unterschiedet man zwischen der Permutation, Variation und Kombination und in diesen 3 Gruppen wird nochmals unterschieden ob Wiederholungen auftreten können oder nicht.

4.2 Permutation

Eine Anordnung von n unterscheidbaren Objekten (es können nicht zwei oder mehr gleiche Objekte auftreten) heißt Permutation ohne Wiederholung (P_n)

$$P_n = n! (1)$$

Wobei n die Anzahl der ausgewählten Objekte ist.

Bei der Permutation mit Wiederholung können verschiedene Sorten auftreten (Sorten sind Mengen mit gleichen Objekten). Für n Objekte und k Sorten gilt:

$$n=n_1+n_2+\ldots+n_k$$
 | n_1,n_2 , etc sind die Sorten
$$P_n^{(n_1,n_2,\ldots,n_k)}=\frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_k!}$$

4.3 Variation

Die Variation is eine Anordnung von k Objekten aus einer Menge aus n Objekten. Wenn ein Objekt mehrfach vorkommen kann spricht man von Variation mit Wiederholung $(V_{w_n}^{(k)})$ ansonsten spricht man von einer Variation ohne Wiederholung $(V_n^{(k)})$. Es gilt:

$$V_n^{(k)} = n(n-1)(n-2)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

 $V_{w_n}^{(k)} = n^k$

4.4 Kombinationen

Bei der Kombination ist im Vergleich zu Variation die Reihenfolge der Objekte nicht wichtig. Sind alle Objekte verschieden spricht man von einer Kombination ohne Wiederholung $(C_n^{(k)})$. Es gilt dann:

$$C_n^{(k)} = \binom{n}{k} = \frac{V_n^{(k)}}{k!}$$

Wenn aus einer Menge von Objekten n Sorten existieren und k Objekte ausgewählt werden und man dazu noch die Reihenfolge nicht berücksichtigt spricht man von einer Kombination mit Wiederholung $(C_{w_n}^{(k)})$. Es gilt:

$$C_{w_n}^{(k)} = \frac{(n-1+k)!}{k!(n-1)!} = \binom{n-1+k}{k}$$

5 Wahrscheinlichkeitstheorie

5.1 Zufallsexperimente

Ein Zufallsexperimente ist ein bleiebig wiederholbarer Verusch, bei dem eine bestimmte Menge als Ausgang möglich ist, die auch bei jeder Wiederholung gleich bleiben muss. Der Ausgang kann dabei auch nicht vohergesagt werden.

- \bullet Jeder einezelene Ausgang heißt Ergebnis (oa. Elemtareignis) und wird mit ω bezeichnet.
- $\bullet\,$ Die Menge aller Ereignise heißt Ω
- ullet Eine Teilmenge von Ω heißt Ereignis
- Ist A eine Teilmenge von Ω so ist \overline{A} das komplementäre Ereignis zu A. \overline{A} beinhält alle Ergebnisse die nicht in A vorkommen $\Rightarrow \overline{A} = \Omega \setminus A$

TODO: Ereignisalgebra

5.2 Laplace Wahrscheinlichkeit

Zufallsexperimente haben endlich viele Ereignisse. Um jetzt die Wahrscheinlichkeit eines bestimmten Ereignis mit anderen einfacher vergleichen zu können teilt man die Wahrscheinlichkeit A durch die Anzahl der Elemente aus Ω^8 . Die Wahrscheinlichkeit ist jetzt in einem Intervall zwischen 1 und 0. Es gilt:

$$p(A) = \frac{|A|}{n} = \frac{\text{Anzahl der Elemente aus } A}{\text{Anzahl der Elemente aus } \Omega}$$
 (2)

⁸Bei Laplace wird angenommen dass alle Ereignisse die selbe Wahrscheinlichkeit haben

- 5.3 Axiome von Kolmogorow
- 5.4 Bedingte Wahrscheinlichkeit
- 5.5 Formel von Bayes
- 5.6 Unabhängigkeit von Ereignissen

6 Zufallsvariable und statistische Verteilung

- 6.1 Zufallsvariable
- 6.2 Diskrete Zufallsvariable
- 6.3 Diskrete Gleichverteilung und Bernoulli-Verteilung
- 6.4 Binomialverteilung
- 6.5 Geometrische Verteilung
- 6.6 Hypergeometrische Verteilung
- 6.7 Poisson-Verteilung
- 6.8 stetige Zufallsvariable
- 6.9 stetige Gleichverteilung
- 6.10 Normalverteilung
- 6.11 Exponentiallverteilung

7 Induktive Statistik und statistische Tests