

Zeitreihen

Was sind Zeitreihen?

Unter einer Zeitreihe versteht man die Beschreibung des Zustands eines Systems in Abhängigkeit zur Zeit t . Der Zustand wird mithilfe einer Zustandsvariablen (K) ausgedrückt, welche beispielsweise einen Kontostand beschreiben kann. Das System besitzt einen Ausgangszustand $K(t_0)$. Der Zustand kann nicht kontinuierlich angegeben werden, sondern nur zu diskreten Zeitpunkten $t_0, t_1, t_2, \dots, t_i$. Diese Zeitpunkte sind äquidistant, also immer um den selben Zeitschritt voneinander entfernt ($t_{i+1} = t_i + h$). Das System enthält eventuell einen oder mehrere Systemparameter p , die die Eigenschaften des Systems beschreiben und die in Abhängigkeit von der Zeit stehen können. $K(t_0)$ und p sind durch Umgebungsbedingungen gegeben oder gegebenenfalls frei wählbar.

Der Zustand zu einem gegebenen Zeitpunkt wird mit der Formel $K(t_{i+1}) = K(t_i) + \Delta K(t_i)$ errechnet, wobei $\Delta K(t_i)$ die Änderung zum Zeitpunkt t_i angibt, das heißt den Wert, um den sich der Zustand nach dem Zeitschritt i ändert. Je nach Wahl der Funktion für $\Delta K(t_i)$ können hier viele verschiedene Systeme abgebildet werden.

$\Delta K(t_i)$ könnte beispielsweise wie folgt aussehen: $\Delta K(t_i) = p \cdot K(t_i)$, wodurch ein einfaches Zinswachstum mit dem Zinssatz p ausgedrückt wird.

Die Funktion $\Delta K(t_i)$ wird in der Literatur auch oft als „Rechte-Seite-Funktion“ bezeichnet.

Zeitabhängigkeit

Die Zeitdifferenz h , also die Länge eines Zeitschritts, spielt in dieser Berechnung eine große Rolle. In der obigen Formel wird sie jedoch nicht berücksichtigt, was keine Variation der Beobachtungsfrequenz zulässt.

Die Formel kann folgenderweise umformuliert werden, um dies einzuschließen:

$$K(t_{i+1}) = K(t_i) + \Delta K(t_i), \text{ wobei } \Delta K(t_i) = h \cdot p \cdot K(t_i);$$

oder auch $K(t_{i+1}) = K(t_i) + h \cdot f(t_i, K(t_i))$

Es gilt: $t_i + h = t_{i+1}$, in den obigen Fällen kann h als 1 festgelegt werden.

Zu bemerken sei hier, dass die Veränderung des Beobachtungsrhythmus eine Veränderung der Ergebnisse hervorrufen kann. Man halte sich das Beispiel des Bankkontos mit Zinsen vor Augen: man bekommt mehr Zinseszins, wenn die Zinsen einmal täglich berechnet werden statt einmal jährlich, auch wenn ihr Prozentsatz nur ein Dreihundertfünfundsechzigstel des Jahreszinses ist. Im zweiten Zeitschritt werden schon die Zinsen aus dem ersten Zeitschritt wieder verzinst, und so fort.

Beispiele

Konto mit Zinsen

Ein einfaches Beispiel für Zeitreihen ist die Kapitalentwicklung eines Bankkontos mit einem festen Zinssatz.

Hier entspricht der Zustand K dem aktuellen Kontostand, $K(t)$ dem Kontostand zum Zeitpunkt t . $K(t_0)$ ist gegeben als der Kontostand zu Beginn der Zeitreihe, also ein bestimmtes Kapital, das einmalig eingezahlt wird.

Die Funktion $\Delta K(t_i)$ lautet wie folgt:

$$\Delta K(t_i) = p \cdot K(t_i)$$

Der Systemparameter p gibt den Zinssatz an.

Dieses Beispiel ist gut als Zeitreihe auszudrücken, weil Zinsen von Banken gewöhnlich in jährlichen Abständen, also zu diskreten, äquidistanten Zeitpunkten ausgezahlt werden, und weil die Kapitalentwicklung durch Zinsen den Gesetzen des Wachstums folgt.

In der praktischen Umsetzung des Modells hat sich gezeigt, dass die Kurve einer exponentiellen Wachstumskurve gleicht.

Entwicklung einer Epidemie

Epidemien entwickeln sich – mathematisch gesehen – nach dem folgenden Prinzip: Eine gewisse Anzahl an Individuen ist infiziert, jeder von ihnen trifft pro Tag eine Anzahl anderer Individuen und steckt sie an.

Nach einer gewissen Zeit jedoch sind schon sehr viele Individuen infiziert, und man kann davon ausgehen, dass jedes von ihnen auch bereits infizierte Individuen trifft. Somit infiziert es nicht alle Individuen, die es trifft, sondern nur einen bestimmten Prozentsatz.

Festzulegen sei hier, dass $K(t_0)$ dem Wert der anfangs infizierten Personen entspricht, und p den Anteil an Individuen von der Gesamtzahl, die jedes einzelne Individuum pro Zeitschritt ansteckt.

Hierbei treten die Regeln des logistischen Wachstums in Kraft. Dies wird in der Funktion $\Delta K(t_i)$ ausgedrückt:

$$\Delta K(t_i) = p \cdot K(t_i) \cdot (L - K(t_i))$$

L entspricht der Gesamtzahl aller Individuen (also sowohl infizierte als auch gesunde). Somit entspricht $L - K(t_i)$ der Anzahl gesunder Individuen. Also wird der Prozentsatz p zunächst mit der Anzahl infizierter Personen multipliziert, und danach mit der Anzahl gesunder Individuen, was dazu führt, dass die Wachstumsrate geringer wird, je mehr Individuen schon infiziert sind.

Man beachte, dass p ein sehr kleiner Wert ist. Beispielhaft soll nun angenommen werden, dass die Population 100 000 Individuen beträgt und jedes infizierte Individuum zwei weitere ansteckt.

Hierbei ist $p = \frac{2}{100\,000}$, also 0,00002.

Entwicklung der Weltbevölkerung

Ein ähnliches Modell wird verwendet, um die Entwicklung der Weltbevölkerung zu beschreiben oder vorausszusagen. $K(t_0)$ ist die Bevölkerung zu Beginn der Zeitreihe, p der Prozentsatz, um den die Bevölkerung pro Zeitschritt im Durchschnitt wächst, und L die Anzahl Menschen, die schätzungsweise auf der Erde leben können, ohne ihr eigenes Überleben gegenseitig zu behindern.

Es gilt wieder $\Delta K(t_i) = p \cdot K(t_i) \cdot (L - K(t_i))$.

Da die Entwicklung der Weltbevölkerung (Geburten- und Sterberate) prozentual nahezu konstant ist, kann p konstant gewählt werden. Zu beachten ist hier, dass L sich verändern kann. Wie schon beschrieben, gibt es die Anzahl Menschen an, die auf der Erde leben können, ohne sich gegenseitig am Überleben zu hindern. Dies schließt z. B. mit ein, dass genügend Lebensmittel für jeden zur Verfügung stehen, oder der Wohnkomfort im Mittel so hoch ist, dass die verfügbare Wohnfläche auf der Erde ausreicht. Man halte sich nun beispielsweise die Entdeckung Amerikas vor Augen, welches vorher sehr dünn besiedelt war, wohingegen Europa einen relativ dichten Besiedelungsquotienten hatte. Nun entdeckten die Europäer Amerika, was bedeutete, dass sie wieder eine sehr große Menge an Land zum Bewohnen und Bestellen zur Verfügung hatten. Somit hat sich im Modell L erhöht.

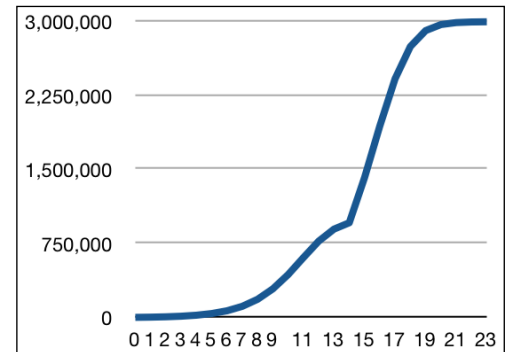


Abb. 1: Entwicklung der Weltbevölkerung (beispielhaft)

Im Schaubild kann man hier einen Knick in der Kurve erkennen. (Siehe Abb. 1)

Im Modell wurde für $\Delta K(t_i)$ die Formel

$$\Delta K(t_i) = p \cdot K(t_i) \cdot \frac{(L - K(t_i))}{L}$$

gewählt, um im letzten Faktor nicht eine absolute Zahl, sondern einen Anteil zu erhalten. Dies dient nur Demonstrationszwecken, da ansonsten die Zahlen zu groß und der „Knick“ zu undeutlich würden.

Veränderung von h

In diesem Beispiel nimmt man an, dass sich der Bestand K in jedem Zeitschritt (d. h. jedes Mal, wenn t um 1 erhöht wird) verdoppelt wird. Es könnte sich hierbei etwa um eine Zellpopulation handeln.

Gewählt wurde die Formel $K(t_{i+1}) = K(t_i) + h \cdot \Delta K(t_i)$, wobei $\Delta K(t_i) = 1 \cdot K(t_i)$.

Im Schaubild ist gut zu erkennen, dass bei einer Messung (und damit auch einer Berechnung von ΔK) alle 0,5 Zeiteinheiten schon nach 10 Zeitschritten etwa der ein Wert gemessen wird, der dreimal so hoch ist wie bei einer Messung mit $h = 1$. Dies ist der sogenannte Zinseszinsseffekt.

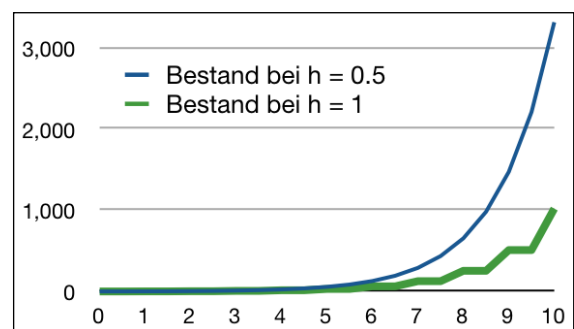


Abb. 2: Veränderung des Ergebnisses bei Veränderung von h

Räuber-Beute-Modelle (nach Lotka und Volterra)

Alfred James Lotka und Vito Volterra haben 1925 und 1926 unabhängig voneinander Regeln aufgestellt, nach welchen Gesetzmäßigkeiten sich Räuber-Beute-Populationen verhalten. Eine dieser Regeln besagt, dass sowohl die Veränderung der Räuberpopulation als auch die Veränderung der Beutepopulation von einander abhängen.

Lotka und Volterra haben zu diesem Modell diverse Gleichungen entwickelt. Eine hiervon soll nun in vereinfachter Form genutzt werden, um solche Modelle in einer Zeitreihe auszudrücken. Die Darstellung des Zustandes von Räubern und Beute kann realisiert werden, in-

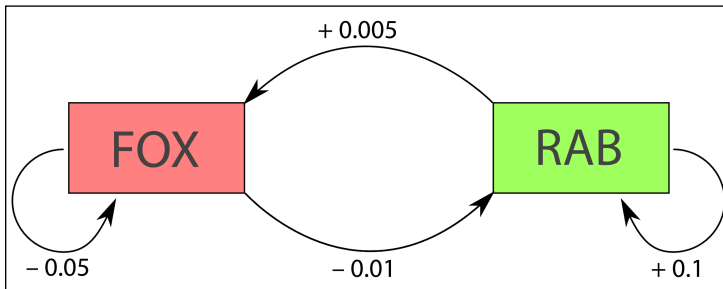


Abb. 3: Visualisierung eines einfachen Räuber-Beute-Modells. Gegeben sind Füchse (FOX) als Räuber und Hasen (RAB) als Beute.

dem man den Bestand K als Vektor definiert. Er soll zur besseren Unterscheidung nicht mehr K , sondern x genannt werden. Die Delta-Funktion ΔK wird nun einfach als $f(x, t)$ definiert. Damit ergibt sich folgende Gleichung:

$$x(t_{i+1}) = x(t_i) + h \cdot f(x, t_i)$$

Nun soll anhand der obigen Abbildung, die zuvor als Modell gezeichnet wurde, eine Zeitreihe erstellt werden.

x sei hier folgendermaßen definiert:

$x = \begin{pmatrix} F \\ H \end{pmatrix}$, wobei F die Anzahl Füchse (also Räuber) und H die Anzahl Hasen (also Beute) darstellen.

Um nun $f(x, t)$ aufzubauen, werden zunächst die Delta-Funktionen von H und F einzeln definiert:

$$f_H(F, H, t) = -0,01 \cdot F + 0,01 \cdot H$$

$$f_F(F, H, t) = -0,05 \cdot F + 0,005 \cdot H$$

Fasst man diese in Matrixschreibweise zusammen, ergibt sich folgendes $f(x, t)$:

$$f(x, t) = \begin{pmatrix} -0,05 & 0,005 \\ -0,01 & 0,1 \end{pmatrix} \cdot x$$

Dieses System ist noch nicht naturgetreu, um die Beziehung der Hasen und Füchse genauer darzustellen, sollte noch eine Schranke für die Hasenpopulation eingebaut werden. Es stellt jedoch die eigentliche Räuber-Beute-Beziehung bereits recht gut dar.

Räuber-Beute-Modell mit Michaelis-Menten-Term

Die Michaelis-Menten-Theorie (benannt nach ihren Entdeckern, Leonor Michaelis und Maud Menten) bezieht sich eigentlich auf die Reaktion von Enzymen in Substraten. Sie ist jedoch durch Veränderung der Parameter auch auf ein Räuber-Beute-Modell anwendbar.

Schreibt man beispielsweise f_H (die Rechte-Seite-Funktion der Hasenpopulation) wie folgt um, so enthält sie eine Sättigungskonstante L , die das Wachstum der Hasen ähnlich dem logistischen Wachstum beschränkt:

$$f_H(F, H, t) = -0,01 \cdot F + 0,01 \cdot (L - H)$$

Rechnet man dieses Modell nun mit einem Anfangswert von zehn Füchsen und zwei Hasen durch, so pendelt sich die Population nach einigen Zeitschritten auf einen konstanten Wert ein, wie in Abbildung 4 zu sehen.

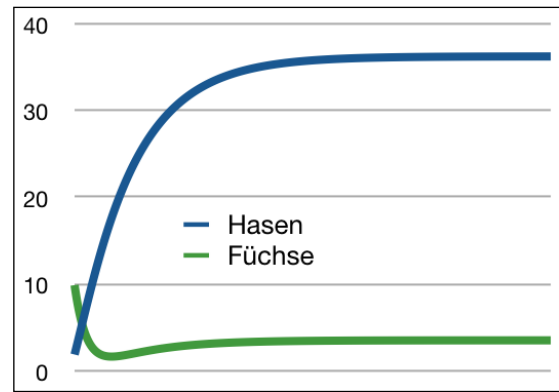


Abb. 4: Populationsentwicklung von Räubern und Beute

Programmiertechnische Umsetzung

Ansatz und Beschreibung

Der Grundansatz ist es, eine Klasse *TimeSeries* für die generelle Berechnung von Zeitreihen anzulegen. In ihr werden alle Werte für K als Array abgelegt, sie kümmert sich auch um die Speicherverwaltung hierfür.

Diese Klasse berechnet allerdings nicht ΔK – hierfür wurde eine eigene Klasse *Delta* mit einer virtuellen Methode zum Berechnen von ΔK anhand von t und K angelegt. *Delta* selbst berechnet jedoch nichts, hierfür muss eine Klasse erstellt werden, die von *Delta* erbt, und in der die vormals virtuelle Methode ΔK berechnet und zurückgibt.

Da dieses Problem durch Klassenvererbung angegangen wurde, kann der Programmierer der Zeitreihe selbst Parameter geben, indem er sie einfach in der von *Delta* erbenden Klasse definiert.

Die Klasse *TimeSeries* berechnet nun anhand eines übergebenen Startwertes und *Delta* $\Delta K(t_0)$, addiert gemäß der oben aufgestellten Formel

$$K(t_{i+1}) = K(t_i) + \Delta K(t_i)$$

$K(t_0)$ und speichert das Ergebnis als K_{t_i} , und so fort.

In beiden Klassen wurden C++-Templates verwendet, was bedeutet, dass die Klassen für sämtliche Datentypen verwendbar sind, also auch eigene Klassen, die dynamische Systeme oder Ähnliches beschreiben. Sie müssen lediglich die Operatoren „+“ und „*“ überladen oder unterstützen.

Anwendung

Zunächst wird eine von der Klasse *Delta* abgeleitete Klasse geschrieben. Diese muss die Methode *calculate* enthalten, deren Deklaration wie folgt aussieht:

```
virtual Object calculate (double t, Object *stock)
```

wobei *Object* ein Klassentemplate ist. *stock* entspricht K .

In dieser Methode wird ΔK berechnet und zurückgegeben.

Nun erstellt man ein neues *TimeSeries*-Objekt, dem als Templateargument der gewünschte Datentyp (*int*, *double*, eigene Klassen, etc.) übergeben wird. Als Argumente nimmt der Konstruktor einen Startwert (K_{t_0}) und einen Zeiger auf die neu erstellte *Delta*-Klasse entgegen.

`TimeSeries::next()` berechnet den K -Wert für den nächsten Zeitschritt und gibt ihn zurück.

`TimeSeries::calculateSteps(double n)` berechnet die nächsten n Schritte und gibt die Werte per `TimeSeries::printStep(double t, Object result)` aus. Dies ist eine virtuelle Funktion und kann so durch Vererbung ersetzt werden, um die Ausgabe abzuändern oder die Daten anderweitig zu verändern.

t kann beiden Klassen als Fließkommazahl übergeben werden. Gespeichert werden nur K -Werte zu ganzzahligen t -Werten, jedoch bei Übergabe und Berechnung wird der genaue Wert weiterverwendet oder neu berechnet.

Von der Zeitreihe zur gewöhnlichen Differentialgleichung

Zeitreihen, wie sie hier vorgestellt wurden, können auch in Differentialgleichungen ausgedrückt werden. Eine Differentialgleichung beschreibt das Verhältnis einer (gewöhnlicherweise unbekannt) Funktion zu ihrer bekannten Ableitung. Beispielsweise kennen wir beim obigen Bankkonten-Beispiel nicht die Funktion, die die Kurve erzeugt, sondern nur ihre Ableitung (hier ΔK , in der Differentialrechnung meist f' oder \dot{f}) und einen gegebenen Anfangswert, z. B. $K(t_i) = 1000$.

Als Differentialgleichung würde man dies so schreiben:

$$\dot{K} = p \cdot K$$

Aus der Mathematik ist bekannt, dass die Ableitung einer Funktion die Steigung der Funktion zum jeweiligen Argumentwert angibt. Die Steigung nach der Stelle $K(t_i)$ ist also $p \cdot K(t_i)$, also $p \cdot 1000 - p$ ist ein Parameter und damit bekannt. Der Funktionswert von $K(t_{i+1})$ ist demnach der alte Funktionswert plus die Steigung, was in folgender Formel resultiert:

$$K(t_{i+1}) = K(t_i) + \dot{K}(t_i)$$

Dies entspricht genau der Formel in der obigen Beschreibung.

Dieses Lösungsverfahren wird als „eulersches Polygonzugverfahren“, in der Literatur auch als „explizites Eulerverfahren“ bezeichnet und ist eine einfache, aber auch recht ungenaue Möglichkeit, numerische Funktionswerte von Differentialgleichungen zu berechnen.

Bei dieser Methode ist der gemachte Fehler für jeden Zeitschritt umso höher, je größer der Zeitschritt ist. Die höchste Genauigkeit, die man erzielen kann, entspricht etwa der Hälfte der berechneten Nachkommastellen.