6. Unabhängige Zufallsvariablen

Diskrete Stochastik

Prof. Dr. Andreas Vogt

Fachhochschule Nordwestschweiz

Unabhängigkeit von Zufallsvariablen

Definition:

Eine unendliche Folge von Zufallsvariablen heisst (stochastisch) unabhängig, wenn jede endliche Teilfolge davon stochatisch unabhängig ist.

Eine endliche Folge von Zufallsvariablen $X_1, \ldots, X_n$ heisst (stochastisch) unabhängig, wenn

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdot \ldots \cdot P(X_n \in A_n)$$

für alle Teilmengen $A_1, A_2, \ldots, A_n$ von $\mathbb{R}$ gilt.

Beispiel: Wir würfeln mit einem fairen Würfel dreimal.

Die Zufallsvariable X zähle die Anzahl an gewürfelten Einsen.

Die Zufallsvariable Y zähle die Anzahl an gewürfelten Vieren in den ersten beiden Würfen.

Dann sind X und Y nicht stochastisch unabhängig, da $P(X = 3, Y = 2) = 0 \neq \left(\frac{1}{6}\right)^3 \cdot \left(\frac{1}{6}\right)^2 = P(X = 3) \cdot P(Y = 2)$.

Unabhängigkeit von Zufallsvariablen

Howard Wolowitz: [after everyone cheers for him and his team design going to space] It gets better! Someone has to go up with the telescope as a payload specialist, and guess who that someone is!

Sheldon Cooper: Mohammed Lee. *[everyone's looking confused]*

Howard Wolowitz: Who's Mohammed Lee?

Sheldon Cooper: Mohammed is the most common first name in the world, and Lee the most common surname. As I didn't know the answer, I thought that'd give me a mathematical edge.

Es bezeichne X bzw. Y die Zufallsvariable, die den Vornamen bzw. den Nachnamen einer Person angibt, wenn man zufällig eine Person der Weltbevölkerung wählt.

Dann ist $P(X = Mohammed)$ maximal, (d.h. grösser als $P(X = x)$ für alle anderen Vornamen x).

Ebenso ist $P(Y = Lee)$ maximal.

Wenn nun X und Y unabhängig wären, dann wäre $P(X = Mohammed, Y = Lee) = P(X = Mohammed) \cdot P(Y = Lee)$, also wäre $P(X = Mohammed, Y = Lee)$ grösser als $P(X = x, Y = y)$ für alle anderen x, y.

Offensichtlich sind X und Y nicht unabhängig, tatsächlich ist die Kombination *Mohammed Lee* wohl ziemlich unwahrscheinlich.

Unabhängigkeit von Zufallsvariablen

Wir modellieren oft definitionsgemäss unabhängige Vorgänge mittels verschiedenen Zufallsvariablen.

Beispiel:

Person 1 kommt zu einem zufälligen Zeitpunkt zwischen 12:00 und 12:45 Uhr, Person 2 unabhängig davon zwischen 12:15 und 13:00 Uhr in ein Café.

Wie hoch ist die Wahrscheinlichkeit, dass beide vor 12:30 Uhr ankommen?

X: Ankunftszeit von Person 1 $\quad X \sim U[0, 45]$

Y: Ankunftszeit von Person 2 $\quad Y \sim U[15, 60]$

In diesem Modul verzichten wir auf eine genaue Konstruktion von Ergebnisräumen für die Zufallsvariablen und entsprechenden Wahrscheinlichkeitsmassen, so dass die Zufallsvariablen, die in der Definition von "Unabhängigkeit" geforderte Bedingung erfüllen.

Wir sehen ein geeignetes Wahrscheinlichkeitsmass P als gegeben an und verwenden die Unabhängigkeitsbedingung dafür:

$$P(X \leq 30, Y \leq 30) = P(X \leq 30) \cdot P(Y \leq 30) = \frac{30}{45} \cdot \frac{15}{45} = \frac{2}{9}.$$

Unabhängigkeit von Zufallsvariablen

In vielen Bereichen spielen Summen von (unabhängigen) Zufallsvariablen eine wichtige Rolle.

Dabei möchte man gerne Aussagen über die Verteilung der Summe von Zufallsvariablen machen.

Beispiel:

Wir werfen solange eine Münze, bis zum ersten Mal Kopf kommt, dann würfeln wir solange, bis eine 6 kommt.

Wie hoch ist die Wahrscheinlichkeit, dass wir z Schritte warten müssen?

X: Anzahl Münzwürfe Y: Anzahl Würfelwürfe

Gesucht: Verteilung von $X + Y$

Modell: $X \sim Geo\left(\frac{1}{2}\right)$; $Y \sim Geo\left(\frac{1}{6}\right)$; X, Y unabhängig

$$P(X + Y = z) = \sum_{x=1}^{z} P(X = x, Y = z - x) = \sum_{x=1}^{z} P(X = x) \cdot P(Y = z - x) =$$

$$= \sum_{x=1}^{z} geopdf(x - 1, \frac{1}{2}) \cdot geopdf(z - x - 1, \frac{1}{6})$$

Allgemein erhält man analog folgendes Ergebnis:

Satz:

Es seien X und Y unabhängige diskrete Zufallsvariablen mit Werten in $\mathcal{X}$ bzw. $\mathcal{Y}$ und Zähldichten f_X bzw. f_Y. Dann hat die Summe $X + Y$ die Zähldichte

$$f_{X+Y}(z) = \sum_{x \in \mathcal{X}} f_X(x) \cdot f_Y(z - x).$$

Hierbei ist $f_Y(y) = 0$ für $y \notin \mathcal{Y}$.

Die Zähldichte f_{X+Y} nennt man auch Faltung von f_X und f_Y.

Mit obiger Formel kann man etwa folgendes nachrechnen:

Wenn X und Y unabhängige Zufallsvariablen sind, dann gilt:

(1) $X \sim Poi(\lambda_1)$, $Y \sim Poi(\lambda_2) \Rightarrow X + Y \sim Poi(\lambda_1 + \lambda_2)$

(2) $X \sim Bin(n_1, p)$, $Y \sim Bin(n_2, p) \Rightarrow X + Y \sim Bin(n_1 + n_2, p)$

Beispiel:

Im Schnitt ereignen sich in einem bestimmten Land 3 Taifune und 2 Erdbeben pro Jahr. Man geht davon aus, dass diese Katastrophen unabhängig voneinander sind. Wie hoch ist die Wahrscheinlichkeit für mehr als 2 Katastrophen im Jahr?

X: Anzahl Taifune, Y: Anzahl Erdbeben

$X \sim Poi(3)$, $Y \sim Poi(2) \quad \Rightarrow X + Y \sim Poi(5) \quad \Rightarrow P(X + Y > 2) = 1 - \texttt{poisscdf}(2, 5)$

Unabhängigkeit von Zufallsvariablen

Für stetig verteilte Zufallsvariablen gilt folgender Satz:

Satz:

Es seien X und Y unabhängige stetig verteilte Zufallsvariablen mit Dichten f_X bzw. f_Y. Dann hat die Summe $X+Y$ die Dichte

$$f_{X+Y}(z) = \int_{x=-\infty}^{\infty} f_Y(z-x) f_X(x) dx.$$

Mit obiger Formel kann man etwa folgendes nachrechnen:

Additonstheorem der Normalverteilung:

Es seien $X_1, X_2, \ldots, X_n$ unabhängige, normal verteilte Zufallsvariabeln eines Zufallsexperimentes mit Erwartungswerten μ_i und Standardabweichungen σ_i. Weiter seien $a_1, a_2, \ldots, a_n$ beliebige reelle Zahlen, nicht alle gleich 0.

Dann ist die Zufallsvariable

$$Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

auch normal verteilt mit dem Erwartungswert $a_1\mu_1 + \ldots + a_n\mu_n$ und der Varianz $a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2$.

Unabhängigkeit von Zufallsvariablen

Beispiel:

Konservendosen mit Trockenmilch werden in Kartons à 50 Dosen verpackt.

Es sei Gewicht des Inhaltes einer Dose, $\mathcal{N}(250g, 7.5g)$ verteilt

Leergewicht einer Dose, $\mathcal{N}(30g, 2.2g)$ verteilt

Leergewicht eines Kartons, $\mathcal{N}(750g, 20g)$ verteilt

Mit welcher Wahrscheinlichkeit wiegt ein Karton mehr als $15kg$?

Das Gesamtgewicht ist dann $\mathcal{N}(\underbrace{50{\cdot}250g+50{\cdot}30g+750g}_{=14750\text{g}}, \underbrace{\sqrt{50\cdot 7.5^2+50\cdot 2.2^2+20^2}g}_{=58.775g})$ verteilt.

Also: $P(\text{Gesamtgewicht} \geq 15000) = 1 - normcdf(15000, 14750, 58.775) \approx 0.001\%$

Additionstheorem Normalverteilung

Eine gewisse Maschine produziert Holzleisten, deren Länge normalverteilt mit $\mu = 100cm$ und $\sigma = 5cm$ ist. Sie benötigen zwei Leisten, die zusammen nicht länger als $202cm$ sein dürfen.

Welche Variante ist besser?

Variante A:

Sie lassen sich zwei Leisten von der Maschine produzieren. (Die Leisten werden unabhängig voneinander produziert.)

Variante B:

Sie lassen sich eine Leiste von der Maschine produzieren und sägen selber aus einer gegeben (grösseren) Leiste eine exakt gleich grosse Leiste zu.

$$X_1 + X_2 \sim \mathcal{N}(100 + 100, \sqrt{2} \cdot 5)$$

$$1 - normcdf(202, 200, sqrt(2) * 5) = 0.3886$$

$$2X \sim \mathcal{N}(2 \cdot 100, 2 \cdot 5)$$

$$1 - normcdf(202, 200, 2 * 5) = 0.4207$$

Serie 6, A.1–A.6

Die Ungleichung von Tschebyscheff

Satz: Es sei X eine Zufallsvariable mit Erwartungswert μ und endlicher Varianz σ^2.

Dann gilt für alle $k > 0$: $P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$.

Diese Abschätzung gilt für alle möglichen Verteilungen von X. Sie ist deshalb in manchen Fällen recht grob.

Beispiel:
Wir nehmen an, dass die Grösse einer erwachsenen Person einen Erwartungswert von $175cm$ bei einer Standardabweichung von $10cm$ hat.

(1) Schätzen Sie mit der Formel von Tschebyscheff die Wahrscheinlichkeit dafür ab, dass eine Person kleiner als 160cm oder grösser als 190cm ist.
(2) Vergleichen Sie diese Abschätzung mit dem exakten Wert unter einer Normalverteilungsannahme.

Lösung: (1) $P(|X - 175cm| \geq 15) \leq \frac{100}{15^2} \approx 44.4\%$.

(2) $P(|X - 175cm| \geq 15) = 1 - (normcdf(190, 175, 10) - normcdf(160, 175, 10)) \approx 13.4\%$.

Aus (2) von Folie 6.5 folgt, dass $\sum_{i=1}^{n} X_i \sim Bin(n,p)$ für eine Folge unabhäniger $B(p)$-verteilter Zufallsvariablen X_i.

Aus der Tschebyscheff-Ungleichung erhalten wir:

$\boxed{P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}}$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| \geq \epsilon\right) = P\left(\left|\sum_{i=1}^{n} X_i - np\right| \geq n\epsilon\right) \leq \frac{n \cdot p \cdot (1-p)}{n^2\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \to 0 \ (n \to \infty).$$

Interpretation:

Die Wahrscheinlichkeit, dass die relative Häufigkeit der Treffer um mehr als ϵ von der theoretischen Wahrscheinlichkeit für Treffer abweicht, geht gegen 0.

Dies gilt nicht nur bei Binomial-verteilten Zufallsvariablen, sondern auch in vielen anderen Situationen:

(Schwaches) Gesetz der grossen Zahlen:

Es sei E ein Ereignis eines Zufallsexperimentes, und sei p die Wahrscheinlichkeit von E. Weiter bezeichne X die Anzahl des Eintreffens von E bei n unabhängigen Ausführungen dieses Experimentes. Für jede beliebig kleine reelle Zahl $\epsilon > 0$ gilt dann:

$$P\left(\left|\frac{X}{n} - p\right| \geq \epsilon\right) \to 0 \ (n \to \infty).$$

Grenzwertsätze

(1) Stellen Sie die Zähldichten für die Anzahl an gewürfelten Sechsen bei $n = 1, 5, 10, 50, 100$ unabhängigen Würfelwürfen dar.

(2) Vergleichen Sie diese Zähldichten mit der Dichte von $\mathcal{N}(\frac{n}{6}, \sqrt{\frac{5}{36}} \cdot \sqrt{n})$. $\rightarrow$ Matlab

Die Dichten nähern sich immer weiter an!

Auch dies ist ein allgemeines Prinzip. Dies ist Inhalt des zentralen Grenzwertsatzes:

Zentraler Grenzwertsatz:

Es sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariabeln eines Wahrscheinlichkeitsraumes, welche alle dieselbe Verteilung mit Erwartungswert μ und Varianz σ^2 haben. Dann gilt für grosse n:

Die Summe

$$S_n = X_1 + \ldots + X_n$$

besitzt näherungsweise die Normalverteilung $\mathcal{N}(\mu_n, \sigma_n)$ mit

$$\mu_n = n \cdot \mu \text{ und } \sigma_n = \sqrt{n} \cdot \sigma.$$

Es gilt also näherungsweise $\dfrac{S_n - \mu_n}{\sqrt{n} \cdot \sigma} \sim \mathcal{N}(0, 1)$.

Präzise gilt für alle $z \in \mathbb{R}$: $P\left(\frac{S_n - n\mu}{\sqrt{n} \cdot \sigma} \leq z\right) \rightarrow \Phi(z)$ $(n \rightarrow \infty)$.

Hierbei ist Φ die Verteilungsfunktion von $\mathcal{N}(0, 1)$.

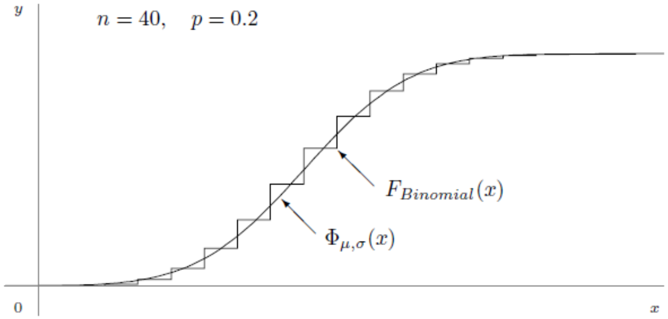
Grenzwertsätze

Da eine $Bin(n,p)$-verteilte Zufallsvariable X als Summe von n $B(p)$-verteilten Zufallsvariablen interpretiert werden kann, gilt insbesondere, dass X näherungsweise $\mathcal{N}(np, \sqrt{np(1-p)})$-verteilt ist:

Grenzwertsatz von de Moivre und Laplace (Spezialfall des zentralen Grenzwertsatzes)

Für grosse n ist

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$



Diese Approximation ist für $n > \frac{9}{p(1-p)}$ hinreichend genau.

Etwas genauer wird es mit der sogenannten Stetigkeitskorrektur:

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

Beispiel:

Ein fairer Würfel wird 1000 mal geworfen. Wie hoch ist die Wahrscheinlichkeit dafür, zwischen 150 und 200 Sechsen zu würfeln?

Präzise: `binocdf(200,1000,1/6)` $-$ `binocdf(149,1000,1/6)` $= 0.9265$

Mit Grenzwertsatz: `normcdf((200+1/2-1000/6)/sqrt(1000*5/36))` $-$ `normcdf((150-1/2-1000/6)/sqrt(1000*5/36))` $= 0.9253$

Simulation von Zufallsvariablen

Manchmal ist es nicht oder nur unter grossen Anstrengungen möglich, gewisse Wahrscheinlichkeiten exakt zu berechnen.

In diesem Fall kann man die Zufallsvariablen simulieren, d.h. Zahlen erzeugen, die entsprechend verteilt sind, und dann schauen, welcher Prozentsatz dieser Zahlen im gesuchten Ereignis liegt.

Gemäss dem Gesetz der grossen Zahlen nähert sich dieser Wert mit wachsender Anzahl an Wiederholungen der tatsächlichen Wahrscheinlichkeit an.

Ziel:
Erzeugung von Zufallszahlen, die gemäss einer gegebenen Verteilung verteilt sind.

Simulation von Zufallsvariablen

In den meisten Programmiersprachen ist ein sogenannter linearer Kongruenzgenerator als Zufallszahlengenerator integriert.

Meist verwendet man die Bezeichnung Pseudozufallszahlengenerator, da die erzeugten Zahlen nicht wirklich zufällig sind, sondern deterministisch, ausgehend von einem Startwert, dem sogenannten Seed, berechnet werden.

in der Praxis oft die Systemzeit

Für kryptographische Zwecke (etwa zur Schlüsselerzeugung) sind lineare Kongruenzgeneratoren nicht geeignet.

Mit nur wenigen Werten kann man die verwendeten Parameter berechnen und so berechnen, welche Zufallszahlen als nächstes konstruiert werden.

Simulation von Zufallsvariablen

Die Zufallszahlenerzeugung mit linearen Kongruenzgeneratoren geht so:

$$x_{n+1} = (a \cdot x_n + b) \bmod m$$

wobei $m > 1$ und $a, b, \overset{\text{Seed}}{\boxed{x_0}} \in \{0, 1 \ldots m-1\}$.

Es wird so sukzessive eine Zahl aus dem (endlichen) Bereich $\{0, 1, \ldots, m-1\}$ berechnet. Der Bereich ist endlich, es entsteht so also eine Periode.

Die Periodenlänge kann dabei sehr kurz sein. Bei $m = 12$, $a = 4$ und $b = 1$ ergibt sich beim Startwert $x_0 = 1$ etwa die Folge: $x_0 = 1$ $x_1 = 5$ $x_2 = 9$ $x_3 = 1$ $x_4 = 5 \ldots$

Der Satz von Knuth besagt, dass die Periodenlänge maximal (also m) ist, wenn $ggT(b, m) = 1$ und wenn $a - 1$ von jedem Primfaktor von m geteilt wird und wenn m durch 4 teilbar ist, dann ist auch $a - 1$ durch 4 teilbar.

Für $m = 8$, $a = 5$, $b = 3$ und $x_0 = 1$ ergibt sich etwa die Folge $1, 0, 3, 2, 5, 4, 7, 6, 1, \ldots$

In `java.util.random` ist etwa $m = 2^{48}$, $a = 25214903917$, $b = 11$, wobei allerdings nur die Bits $47 \ldots 16$ ausgegeben werden.

Durch die Transformation $z_n = a + (b - a)\frac{x_n}{m}$ erhält man eine Zahl im Intervall $[a, b]$.

Simulation von Zufallsvariablen: Inversionsmethode

Wir haben gerade gesehen, wie man Zufallszahlen etwa im Bereich $[0,1]$ erzeugen kann.

Mit der Inversionsmethode ist nun nun möglich, daraus Zufallszahlen zu konstruieren, die gemäss einer Verteilung mit der Verteilungsfunktion F verteilt sind.

Grundlage ist folgender Satz:

Satz:

Es sei F eine streng monoton steigende (also bijektive) Verteilungsfunktion.

Wenn nun U eine im Intervall $[0,1]$ gleichverteilte Zufallsvariable ist, dann ist die Zufallsvariable $F^{-1}(U)$ verteilt mit Verteilungsfunktion F.

Beweis:

$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$, womit die Behauptung gezeigt ist.

Wenn man also gleichverteilte Zufallszahlen im Intervall $[0,1]$ erzeugt und darauf die Funktion F^{-1} anwendet, sind die resultierenden Zahlen gemäss der Verteilungsfunktion F verteilt.

Bemerkung:

Der Satz ist auch dann korrekt, wenn F nicht streng monoton steigend ist. In diesem Fall existiert die Umkehrfunktion F^{-1} nicht. Man verwendet stattdessen die sogenannte Quantilfunktion $F^{-1}(y) = \inf\{x \in \mathbb{R} \mid F(x) \geq y\}$.

Simulation von Zufallsvariablen: Inversionsmethode

Beispiel:

An einer bestimmten Kreuzung kommen im Schnitt λ Autos pro Stunde vorbei. Dann ist die Wartezeit auf das nächste Auto bekanntlich exponentiell verteilt mit Parameter λ.

Wir möchten die Wartezeiten auf jeweils das nächste Auto simulieren.

Die Verteilungsfunktion der Exponentialverteilung lautet: $F(t) = 1 - e^{-\lambda t}, \ (t \geq 0)$.

Die Umkehrfunktion erhalten wir via: $F(t) = y \Leftrightarrow 1 - e^{-\lambda t} = y \Leftrightarrow t = -\frac{\ln(1-y)}{\lambda}$.

Es ergibt sich $F^{-1}(y) = -\frac{\ln(1-y)}{\lambda}$.

```
n=100000;
lambda=5;
u=rand([n,1]);
t=-log(1-u)/lambda*60;
```

simuliert in Matlab etwa die Wartezeiten in Minuten auf n Autos.

```
z=0;
for i=1:n
   if t(i)>5
     z=z+1;
   end
end
z/n
```

gibt den Anteil aller Wartezeiten von mehr als 5 Minuten aus

Die theoretische Wahrscheinlichkeit dafür beträgt bekanntlich

$$P(T > 5min) = 1 - \texttt{expcdf}(1/12, 1/5) = 0.6592.$$

Simulation von Zufallsvariablen

Für normalverteilte Zufallsvariablen gibt es effizientere Simulationsmethoden als die Inversionsmethode, da bei F^{-1} bei der Normalverteilung nicht geschlossen angegeben werden kann.

Bekannte Verfahren sind etwa das Box-Muller-Verfahren oder die Polar-Methode.

Wir können darauf hier nicht eingehen.