KAPITEL 14

# Irrfahrt

## 14.1. Berechnung einer Ruinwahrscheinlichkeit

Das folgende sogenannte *Ruinproblem* wurde (in einer etwas anderen Formulierung) von Pascal in einem Brief an Fermat im Jahr 1656 gestellt.

Beispiel 14.1.1. Man stelle sich folgendes Spiel vor. Am Anfang des Spiels besitzt man ein Startkapital von $k$ Euro, wobei $k \in \mathbb{N}_0$. In jeder Spielrunde kann man entweder mit Wahrscheinlichkeit $p$ ein Euro gewinnen (wodurch sich das Kapital um 1 vergrößert) oder mit Wahrscheinlichkeit $q$ ein Euro verlieren (wodurch sich das Kapital um 1 verkleinert). Hierbei gilt, dass $p, q \in (0, 1)$ und $p + q = 1$. Die Spielrunden seien unabhängig voneinander. Erreicht das Kapital zu einem Zeitpunkt den Wert 0, so sagen wir, dass ein *Ruin* eingetreten ist. Dieses Ereignis bezeichnen wir mit $R$. Nach dem Ruin wird das Spiel beendet. Erreicht das Kapital zu einem Zeitpunkt einen zuvor vorgegebenen Wert $M \geq k$ (der *Zielkapital* genannt wird), so steigt man aus dem Spiel aus. Dieses Ereignis bezeichnen wir mit $G$. Die Spielrunden werden wiederholt, bis eines der beiden Ereignisse $R$ oder $G$ eintritt. Man bestimme die Wahrscheinlichkeiten der Ereignisse $R$ und $G$.

Lösung. Schritt 0. Es sei $p_i$ die Wahrscheinlichkeit von Ruin in einem Spiel, das mit Startkapital $k = i$ startet. In der ersten Spielrunde kann man entweder 1 Euro gewinnen (Wahrscheinlichkeit $= p$) oder 1 Euro verlieren (Wahrscheinlichkeit $= q$). Hat man 1 Euro gewonnen, so hat man nach der ersten Runde ein Kapital von $i+1$ Euro und die Wahrscheinlichkeit, dass irgendwann später der Ruin eintreten wird, ist $p_{i+1}$. Hat man 1 Euro verloren, so hat man nach der ersten Runde ein Kapital von $i - 1$ Euro und die Wahrscheinlichkeit, dass irgendwann später der Ruin eintriten wird, ist gleich $p_{i-1}$. Nach dem Satz der totalen Wahrscheinlichkeit erhalten wir, dass

$$p_i = p \cdot p_{i+1} + q \cdot p_{i-1}, \quad i = 1, \ldots, M-1. \tag{14.1.1}$$

Außerdem gelten trivialerweise die "Randbedingungen" $p_0 = 1$ und $p_M = 0$. Wir haben also ein lineares Gleichungssystem mit $M + 1$ unbekannten, nämlich $p_0, \ldots, p_M$, und $M + 1$ Gleichungen. Dieses Gleichungssystem werden wir nun lösen.

Schritt 1. Es gilt

$$p_i = p \cdot p_i + q \cdot p_i, \quad i = 1, \ldots, M-1, \tag{14.1.2}$$

denn $p + q = 1$. Zieht man nun die Gleichungen (14.1.1) und (14.1.2) voneinander ab, so erhält man

$$p(p_{i+1} - p_i) + q(p_{i-1} - p_i) = 0.$$

Führt man die Notation $d_i := p_i - p_{i+1}$ ein, so erhält man

$$d_i = \frac{q}{p}\, d_{i-1}, \quad i = 1, \dots, M-1.$$

Benutzt man diese Gleichung iterativ, so erhält man

$$d_i = \frac{q}{p}\, d_{i-1} = \left(\frac{q}{p}\right)^2 d_{i-2} = \ldots = \left(\frac{q}{p}\right)^i d_0, \quad i = 1, \dots, M-1. \tag{14.1.3}$$

Schritt 2. Nun berechnen wir $d_0$. Dazu setzen wir die Randbedingungen $p_0 = 1$ und $p_M = 0$ ein:

$$1 = p_0 - p_M = \sum_{i=0}^{M-1} d_i = d_0 \sum_{i=0}^{M-1} \left(\frac{q}{p}\right)^i. \tag{14.1.4}$$

Wir müssen die Summe dieser geometrischen Folge berechnen. Hierzu unterscheiden wir zwei Fälle:

Fall a. Es gelte $p = q = 1/2$, also $q/p = 1$. Dann ergibt sich aus (14.1.4), dass $1 = d_0 \cdot M$, also $d_0 = 1/M$. Mit (14.1.3) folgt, dass $d_i = 1/M$ und somit

$$p_i = p_i - p_M = d_i + \ldots + d_{M-1} = \frac{M-i}{M}, \quad i = 0, \dots, M.$$

Somit ist $p_i$ eine lineare Funktion von $i$, die zwischen den Werten $p_0 = 1$ und $p_M = 0$ interpoliert.

Fall b. Es gelte $p \neq q$, also $r := q/p \neq 1$. Dann ergibt sich aus (14.1.4), dass $1 = d_0 \cdot \frac{1-r^M}{1-r}$. Daraus folgt, dass $d_0 = \frac{1-r}{1-r^M}$. Mit (14.1.3) erhalten wir, dass

$$d_i = r^i d_0 = \frac{r^i - r^{i+1}}{1-r^M}, \quad i = 0, \dots, M-1.$$

Daraus erhalten wir schließlich

$$p_i = p_i - p_M = \sum_{j=i}^{M-1} d_j = \sum_{j=i}^{M-1} \frac{r^i - r^{i+1}}{1-r^M} = \frac{r^i - r^M}{1-r^M}, \quad i = 0, \dots, M.$$

Wir haben also die Ruinwahrscheinlichkeit bestimmt und folgendes Ergebnis erhalten: Bei Startkapital $k$ ist die Ruinwahrscheinlichkeit gleich

$$p_k = \begin{cases} \frac{M-k}{M}, & \text{falls } p = q = \frac{1}{2}, \\ \frac{r^k - r^M}{1-r^M}, & \text{falls } r := \frac{q}{p} \neq 1, \end{cases} \quad k = 0, \dots, M. \tag{14.1.5}$$

Es sei bemerkt, dass der erste Fall als der Grenzwert des zweiten Falls angesehen werden kann, denn

$$\lim_{r \to 1} \frac{r^k - r^M}{1-r^M} = \frac{M-k}{M}.$$

Völlig analog kann man ein Gleichungssystem zur Bestimmumg der Wahrscheinlichkeit des Ereignisses $G$ aufstellen und lösen. Es gibt aber einen schnelleren Weg, die Wahrscheinlichkeit von $G$ zu berechnen.

Schritt 3. Damit das Ereignis $R$ eintritt, muss man $k$ Euro verlieren, bevor man $M-k$ Euro gewonnen hat. Dabei ist die Wahrscheinlichkeit, ein Euro in einer Runde zu verlieren, gleich $q$. Damit das Ereignis $G$ eintritt, muss man $M-k$ Euro gewinnen, bevor man $k$ Euro verloren hat. Dabei ist die Wahrscheinlichkeit, ein Euro in einer Runde zu gewinnen, gleich $p$. Um die Wahrscheinlichkeit von $G$ zu bestimmen, müssen wir also in der Formel für die Wahrscheinlichkeit von $R$ folgende Substitutionen machen:

$$p \leftrightarrow q, \quad r \leftrightarrow 1/r, \quad k \leftrightarrow M-k.$$

Für die Wahrscheinlichkeit $q_k$ von $G$ bei Startkapital $k$ ergibt sich die folgende Formel:

$$q_k = \begin{cases} \frac{M-(M-k)}{M} = \frac{k}{M}, & \text{falls } p = q = \frac{1}{2}, \\ \frac{\frac{1}{r^{M-k}} - \frac{1}{r^M}}{1-\frac{1}{r^M}} = \frac{1-r^k}{1-r^M}, & \text{falls } p \neq q. \end{cases}$$

Schaut man sich nun die Formeln für $p_k$ und $q_k$ an, so erkennt man, dass $p_k+q_k = 1$. Das Spiel wird also mit Wahrscheinlichkeit 1 in einer endlichen Zeit beendet. Die Wahrscheinlichkeit, dass das Spiel ewig dauert, ist somit 0.

Beispiel 14.1.2. Nun nehmen wir an, dass das Zielkapital $M = +\infty$ ist. Man spielt also, bis man ruiniert ist (bzw. ewig, wenn der Ruin nicht eintritt). Wir berechnen die Ruinwahrscheinlichkeit $p_k$ bei Startkapital $k$. Dazu lassen wir $M \to \infty$ in (14.1.5):

$$p_k = \begin{cases} 1, & \text{falls } p = q = 1/2, \\ 1, & \text{falls } p < q, \\ r^k, & \text{falls } p > q. \end{cases}$$

Im Fall $p < q$ ist das Spiel *unvorteilhaft* (man verliert mit einer größeren Wahrscheinlichkeit, als man gewinnt). Deshalb ist es keine Überraschung, dass in diesem Fall der Ruin mit Wahrscheinlichkeit 1 eintritt. Der Ruin tritt aber auch dann mit Wahrscheinlichkeit 1 ein, wenn das Spiel *fair* ist (d.h. im Fall $p = q = 1/2$). Im Fall $p > q$ ist das Spiel *vorteilhaft* (man gewinnt mit einer größeren Wahrscheinlichkeit, als man verliert). In diesem Fall ist die Ruinwahrscheinlichkeit gleich $r^k$. Man beachte, dass bei einem Startkapital, das gegen $+\infty$ strebt, die Ruinwahrscheinlichkeit gegen 0 geht.

## 14.2. Rückkehr der Irrfahrt zum Ursprung

Definition 14.2.1. Seien $X_1, X_2, \ldots$ unabhängige identisch verteilte Zufallsvariablen mit

$$\mathbb{P}[X_i = 1] = p \text{ und } \mathbb{P}[X_i = -1] = q,$$

wobei $p + q = 1$ und $p, q \in (0, 1)$. Dann ist eine *Irrfahrt* die Folge $S_0, S_1, S_2, \ldots$ mit

$$S_0 = 0, \quad S_n = X_1 + \ldots + X_n, \quad n \in \mathbb{N}.$$

Die Irrfahrt heißt *symmetrisch*, wenn $p = q = 1/2$.

Bemerkung 14.2.2. Eine Irrfahrt ist beispielsweise ein Modell für ein Glücksspiel oder ein diffundierendes Teilchens.

Sei nun $Z$ der erste Zeitpunkt, zu dem die Irrfahrt zum Ursprung zurückkehrt, d.h.

$$Z = \min\{n \in \mathbb{N} : S_n = 0\}.$$

Wenn es keine Rückkehr zum Ursprung gibt, so definieren wir $Z = \infty$. Wie ist nun die Zufallsvariable $Z$ verteilt? Wir definieren die Zähldichte von $Z$:

$$f_n = \mathbb{P}[Z = n], \quad n \in \mathbb{N}.$$

Satz 14.2.3. *Für die Zähldichte des ersten Rückkehrzeitpunktes gilt*

$$f_{2n} = \frac{1}{2n-1}\binom{2n}{n} p^n q^n. \qquad f_{2n+1} = 0. \tag{14.2.1}$$

Beweis.

Schritt 1. Unser Ziel ist es, die Zähldichte $f_n$ zu berechnen. Dazu betrachten wir eine verwandte Größe, nämlich

$$u_n = \mathbb{P}[S_n = 0], \quad n \in \mathbb{N}_0.$$

Dabei ist $u_0 = 1$. Hier unterscheiden sich $f_n$ und $u_n$ dadurch, dass bei $f_n$ die *erste* Rückkehr zum Ursprung betrachtet wird, bei $u_n$ hingegen *irgendeine* Rückkehr zum Ursprung.
Trivialerweise ist es nicht möglich, nach einer ungeraden Anzahl von Schritten zum Ursprung zurückzukehren. Daher gilt $u_{2n+1} = 0$. Bei gerader Anzahl von Schritten muss die Irrfahrt genau $n$ Schritte nach oben und $n$ Schritte nach unten machen, um zum Zeitpunkt $2n$ wieder bei 0 anzukommen. Daher gilt (wegen der Binomialverteilung)

$$u_{2n} = \mathbb{P}[S_{2n} = 0] = \binom{2n}{n} p^n q^n, \quad u_{2n+1} = 0. \tag{14.2.2}$$

Schritt 2. Nun stellen wir einen Zusammenhang zwischen $u_n$ und $f_n$ her. Damit die Irrfahrt zum Zeitpunkt $n$ zum Ursprung zurückkehrt, muss die *erste* Rückkehr zum Ursprung zu einem Zeitpunkt $k = 1, \ldots, n$ geschehen. Wenn die Irrfahrt zum Zeitpunkt $k$ zum ersten Mal zum Ursprung zurückkehrt, so ist der Fall für $f_k$ eingetreten. Von hier an kann die Irrfahrt beliebig weiter laufen mit der einzigen Bedingung, dass sie zum Zeitpunkt $n$ wieder zum Ursprung zurückkehrt. Dabei darf sie zwischen $k$ und $n$ zum Ursprung zurückkehren, muss aber nicht. Damit erhalten wir zwei unabhängige Ereignisse $f_n$ und $u_{n-k}$ für die beiden Zeitabschnitte. Für $u_n$ gilt somit die Formel:

$$u_n = \sum_{k=1}^{n} f_k \cdot u_{n-k}. \tag{14.2.3}$$

Diese Gleichung werden wir nun benutzen, um $f_k$ zu bestimmen.

Schritt 3. Die Faltung ist schwierig zu handhaben, daher arbeiten wir stattdessen bevorzugt mit erzeugenden Funktionen $u(s)$ und $f(s)$, die folgendermaßen definiert werden:

$$u(s) = \sum_{n=0}^{\infty} u_n s^n, \quad f(s) = \sum_{n=1}^{\infty} f_n s^n, \quad |s| \leq 1.$$

Beide Funktionen sind wohldefiniert für $|s| \le 1$, denn $0 \le u_n \le 1$ und $0 \le f_n \le 1$. Nun kann man (14.2.3) benutzen, um $u(s)$ in Abhängigkeit von $f(s)$ darzustellen:

$$\begin{aligned}
u(s) &= \sum_{n=0}^{\infty} u_n s^n \\
&= 1 + \sum_{n=1}^{\infty} \left( \sum_{k=1}^{n} f_k \cdot u_{n-k} \right) s^n \\
&= 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} \left(f_k s^k\right) \cdot \left(u_{n-k} s^{n-k}\right) \\
&= 1 + \left( \sum_{n=1}^{\infty} f_k s^k \right) \cdot \left( \sum_{l=0}^{\infty} u_l s^l \right) \\
&= 1 + f(s) \cdot u(s).
\end{aligned}$$

Man sieht, dass wir durch die Anwendung der erzeugenden Funktionen die Faltung in ein Produkt umgewandelt haben. Also gilt die Formel

$$f(s) = \frac{u(s) - 1}{u(s)}, \quad s \in [-1, 1].$$

Schritt 4. Um nun $f(s)$ bestimmen zu können, benötigen wir zunächst $u(s)$:

$$u(s) = \sum_{n=0}^{\infty} u_n s^n = \sum_{n=0}^{\infty} \binom{2n}{n} p^n q^n \cdot s^{2n} = \sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} \left(-4pqs^2\right)^n = \frac{1}{\sqrt{1 - 4pqs^2}}.$$

Hier haben wir verwendet, dass

$$\begin{aligned}
\binom{-\frac{1}{2}}{n} &= \frac{-\frac{1}{2} \cdot \left(-\frac{1}{2} - 1\right)\left(-\frac{1}{2} - 2\right) \cdot \ldots \cdot \left(-\frac{1}{2} - n + 1\right)}{n!} \\
&= (-1)^n \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1)}{2^n n!} \\
&= (-1)^n \frac{\binom{2n}{n}}{4^n}.
\end{aligned}$$

Danach haben wir die Newton–Formel $(1 + x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$ mit $\alpha = -1/2$ benutzt.

Schritt 5. Damit gilt

$$f(s) = \frac{u(s) - 1}{u(s)} = \frac{\frac{1}{\sqrt{1-4pqs^2}} - 1}{\frac{1}{\sqrt{1-4pqs^2}}} = 1 - \sqrt{1 - 4pqs^2}.$$

Um nun die Folge $f_n$ zu berechnen, müssen wir die Funktion $f(s)$ in eine Taylor–Reihe entwickeln. Wir benutzen wieder die Newton–Formel $(1 + x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$, aber diesmal mit $\alpha = 1/2$:

$$f(s) = -\sum_{n=1}^{\infty} \binom{\frac{1}{2}}{n} \left(-4pqs^2\right)^n = \sum_{n=1}^{\infty} \frac{1}{2n - 1} \binom{2n}{n} p^n q^n s^{2n}.$$

Dabei haben wir die Formel $-(-4)^n\binom{\frac{1}{2}}{n} = \frac{1}{2n-1}\binom{2n}{n}$ benutzt (Beweis: Übung). Also gilt

$$f_{2n} = \frac{1}{2n-1}\binom{2n}{n}p^n q^n = \frac{u_{2n}}{2n-1}, \quad f_{2n+1} = 0.$$

Die Formel für $f_{2n+1}$ ist übrigens offensichtlich, denn eine Rückkehr zum Ursprung zu einem ungeraden Zeitpunkt ist nicht möglich. Die Formel für $f_{2n}$ ist aber nicht trivial. □

Beispiel 14.2.4. Wir definieren das Ereignis

$$A = \{\exists n \in \mathbb{N} : S_n = 0\} = \text{“Es gibt eine Rückkehr der Irrfahrt zum Ursprung”}.$$

Nun behaupten wir, dass für die Wahrscheinlichkeit dieses Ereignisses gilt:

$$\mathbb{P}[A] = 1 - |2p-1|.$$

Beweis. Es gibt genau dann eine Rückkehr zum Ursprung, wenn $Z \neq \infty$. Somit gilt

$$\mathbb{P}[A] = \mathbb{P}[Z \neq \infty] = f_1 + f_2 + \ldots = f(1).$$

Mit der Formel für $f(s)$ erhalten wir

$$f(1) = 1 - \sqrt{1-4pq} = 1 - \sqrt{(2p-1)^2} = 1 - |2p-1|.$$

dabei haben wir benutzt, dass $q = 1-p$. □

Bemerkung 14.2.5. Im Fall $p = 1/2$ ist die Wahrscheinlichkeit, dass es eine Rückkehr zum Ursprung gibt, gleich 1. Da die Irrfahrt nach einer Rückkehr zum Ursprung wieder neu (und nabhängig von der Vergangenheit) startet, wird sie mit Wahrscheinlichkeit 1 wieder zum Ursprung zurückkehren, usw. Mit anderen Worten, die Irrfahrt kehrt mit Wahrscheinlichkeit 1 sogar unendlich oft zum Ursprung zurück. Man sagt, dass die Irrfahrt für $p = 1/2$ *rekurrent* ist. Für $p \neq 1/2$ it die Irrfahrt *transient*, d.h. sie kehrt mit Wahrscheinlichkeit 1 nur endlich oft zum Ursprung zurück.

Im Fall $p = 1/2$ findet eine Rückkehr zum Ursprung mit Wahrscheinlichkeit 1 statt. Wie lange muss man im Durchschnitt auf die erste Rückkehr warten? Die Antwort auf diese Frage fällt etwas unerwartet aus.

Satz 14.2.6. *Sei $p = 1/2$ und $Z$ der Zeitpunkt der ersten Rückkehr der Irrfahrt zum Ursprung. Dann gilt $\mathbb{E}Z = +\infty$.*

Beweis. Die erzeugende Funktion von $Z$ ist

$$g_Z(s) = f(s) = 1 - \sqrt{1-s^2}.$$

Daraus ergibt sich dann für den Erwartungswert $\mathbb{E}Z = g_Z'(1) = +\infty$. □

Im nächsten Satz werden wir die Wahrscheinlichkeiten $u_n$ und $f_n$ approximativ (für großes $n$) berechnen. Dazu benötigen wir zunächst folgende Notation:

Definition 14.2.7. Zwei Folgen $a_1, a_2, \ldots$ und $b_1, b_2, \ldots$ heißen *asymptotisch äquivalent*, wenn

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 1.$$

Bezeichnung: $a_n \sim b_n$ für $n \to \infty$.

Satz 14.2.8. *Für $p = 1/2$ gilt*

$$u_{2n} \sim \frac{1}{\sqrt{\pi n}} \text{ und } f_{2n} \sim \frac{1}{2\sqrt{\pi n^3}} \text{ für } n \to \infty.$$

Lösung. Zum Beweis benötigen wir die Stirling–Formel:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \quad n \to \infty.$$

Damit erhalten wir

$$u_{2n} = \binom{2n}{n}\frac{1}{2^{2n}} = \frac{(2n)!}{(n!)^2 \cdot 2^{2n}} \sim \frac{\sqrt{2\pi \cdot 2n} \cdot \left(\frac{2n}{e}\right)^{2n}}{\left(\sqrt{2\pi n}\right)^2 \cdot \left(\frac{n}{e}\right)^{2n} \cdot 2^{2n}} = \frac{1}{\sqrt{\pi n}}.$$

Damit gilt auch $f_{2n} = \frac{u_{2n}}{2n-1} \sim \frac{1}{2n} \cdot \frac{1}{\sqrt{\pi n}} = \frac{1}{2\sqrt{\pi n^3}}$. □

Bemerkung 14.2.9. Mit dem obigen Satz kann man Satz 14.2.6 auf eine andere Weise beweisen. Es gilt nämlich

$$\mathbb{E}Z = \sum_{n=1}^{\infty} f_{2n} \cdot 2n = \infty, \quad \text{da } f_{2n} \cdot 2n \sim \frac{1}{\sqrt{\pi n}} \text{ und } \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty.$$

Bemerkung 14.2.10. Sei $K_{2n}$ eine Zufallsvariable, welche angibt, wie oft die Irrfahrt im Zeitintervall $[1, 2n]$ zum Ursprung zurückkehrt, d.h.

$$K_{2n} = \sum_{k=1}^{n} \mathbb{1}_{S_{2k}=0} = \#\{k : S_{2k} = 0, 1 \le k \le n\}.$$

Wir werden nun den Erwartungswert von $K_{2n}$ bestimmen. Man könnte zunächst vermuten, dass der Erwartungswert für großes $n$ approximativ proportional zu $2n$ sein sollte, da man erwartet, dass bei einem beispielsweise doppelt so großen Zeitintervall auch doppelt so viele Rückkehrzeitpunkte auftauchen. Diese Überlegung ist falsch, wie wir nun rechnerisch zeigen:

$$\mathbb{E}K_{2n} = \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{1}_{S_{2k}=0}\right] = \sum_{k=1}^{n} \mathbb{P}[S_{2k} = 0] \sim \frac{1}{\sqrt{\pi}}\sum_{k=1}^{n}\frac{1}{\sqrt{k}} \sim 2\sqrt{\frac{n}{\pi}}, \quad n \to \infty.$$

Der vorletzte Schritt gilt wegen Satz 14.2.8. Zu zeigen, dass $\sum$ und $\sim$ vertauscht werden können, ist eine nichttriviale Übung. Also ist $\mathbb{E}K_{2n}$ approximativ proportional zu $\sqrt{n}$ und nicht zu $n$.

Dieses Ergebnis kann man sich folgendermaßen veranschaulichen. Wir haben gezeigt, dass die Irrfahrt mit Wahrscheinlichkeit 1 unendlich oft zum Ursprung zurückkehrt. In diesem Sinne ist sie mit einer Sinusfunktion vergleichbar. Jedoch ist bei der Sinusfunktion die Anzahl der Rückkehrpunkte proportional zu $n$. Es ist also besser, die Irrfahrt nicht mit einer normalen Sinusfunktion zu vergleichen, sondern mit einer Sinusfunktion, bei welcher der Abstand zwischen den Rückkehrpunkten immer größer wird.

## 14.3. Verteilung des Maximums der Irrfahrt

Sei $S_0, S_1, S_2, \ldots$ eine Irrfahrt mit $p = 1/2$. Wir bezeichnen mit $M_n$ den maximalen Wert der Irrfahrt auf dem Zeitintervall $[0, n]$, d.h.

$$M_n = \max_{k=0,\ldots,n} S_k.$$

Aus $S_0 = 0$ folgt, dass $M_n \geq 0$. Im nächsten Satz bestimmen wir die asymptotische Verteilung von $M_n$ für $n \to \infty$.

Satz 14.3.1. *Für jedes $x > 0$ gilt:*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{M_n}{\sqrt{n}} \leq x\right] = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

*Mit anderen Worten:*

$$\frac{M_n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} |N|, \text{ wobei } N \sim N(0,1).$$

Beweis. Sei $m \in \mathbb{N}$. Mit dem Spiegelungsprinzip (Bild fehlt) erhalten wir

$$\begin{aligned}
\mathbb{P}[M_n \geq m] &= \mathbb{P}[M_n \geq m, S_n < m] + \mathbb{P}[M_n \geq m, S_n = m] + \mathbb{P}[M_n \geq m, S_n > m] \\
&= \mathbb{P}[S_n > m] + \mathbb{P}[S_n = m] + \mathbb{P}[S_n > m] \\
&= 2\mathbb{P}[S_n > m] + \mathbb{P}[S_n = m] \\
&= 2\mathbb{P}[S_n \geq m] - \mathbb{P}[S_n = m].
\end{aligned} \tag{14.3.1}$$

Sei jetzt $x > 0$. Die Zahl $x\sqrt{n}$ muss nicht ganzzahlig sein. Wir können aber eine ganze Zahl $m_n$ mit $m_n - 1 \leq x\sqrt{n} < m_n$ finden. Dann gilt:

$$\begin{aligned}
\mathbb{P}\left[\frac{M_n}{\sqrt{n}} > x\right] &= \mathbb{P}[M_n \geq m_n] \\
&= 2\mathbb{P}[S_n \geq m] - \mathbb{P}[S_n = m] \\
&= 2\mathbb{P}\left[\frac{S_n}{\sqrt{n}} \geq \frac{m_n}{\sqrt{n}}\right] - \mathbb{P}[S_n = m] \\
&\underset{n\to\infty}{\longrightarrow} 2\mathbb{P}[N \geq x] - 0,
\end{aligned}$$

wobei $N \sim N(0,1)$ und wir den zentralen Grenzwertsatz benutzt haben. Außerdem haben wir benutzt, dass $\lim_{n\to\infty} \mathbb{P}[S_n = m_n] = 0$ ist. Hier ist der Beweis. Der Abstand zwischen $x\sqrt{n}$ und $m_n$ ist höchstens 1. Sei $\varepsilon > 0$ fest. Somit liegt $m_n$ immer zwischen $x\sqrt{n}$ und $(x+\varepsilon)\sqrt{n}$, zumindest wenn $n$ groß genug ist. Also gilt:

$$\mathbb{P}[S_n = m_n] \leq \mathbb{P}\left[x \leq \frac{S_n}{\sqrt{n}} \leq x + \varepsilon\right] \underset{n\to\infty}{\longrightarrow} \Phi(x+\varepsilon) - \Phi(x).$$

Dies gilt für jedes $\varepsilon > 0$. Daher können auch $\varepsilon \downarrow 0$ betrachten. Somit gilt wegen der Stetigkeit von $\Phi$

$$\lim_{n\to\infty} \mathbb{P}[S_n = m_n] = 0.$$

Aus den obigen Überlegungen folgt, dass

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{M_n}{\sqrt{n}} > x\right] = 2\mathbb{P}[N \geq x] = \mathbb{P}[|N| > x].$$

Daraus folgt die Aussage des Satzes. □

## 14.4. Arcussinus–Gesetz

Eine weitere unerwartete Eigenschaft der Irrfahrt ist das sogenannte *Arcussinus–Gesetz*. Sei $S_0, S_1, S_2, \ldots$ eine Irrfahrt mit $p = 1/2$. Wir betrachten die Zufallsvariable $T_{2n}$, welche die Zeit angibt, die die Irrfahrt während des Zeitintervalls $[0, 2n]$ auf der positiven Halbachse verbringt. Das heißt,

$$T_{2n} = \sum_{j=0}^{2n-1} \mathbb{1}_{S_j > 0 \text{ oder } S_{j+1} > 0}.$$

Das Ereignis "$S_j > 0$ oder $S_{j+1} > 0$" tritt genau dann ein, wenn die Irrfahrt die Zeit zwischen $j$ und $j+1$ auf der positiven Halbachse verbringt. Im nächsten Satz berechnen wir die Verteilung von $T_{2n}$.

Satz 14.4.1. *Für die Zähldichte von $T_{2n}$ gilt*

$$p_{2k,2n} \overset{def}{=} \mathbb{P}[T_{2n} = 2k] = u_{2k} \cdot u_{2n-2k}, \quad k = 0, \ldots, n.$$

Beweis. Schritt 0. Sei zuerst $k = n$. Es gilt

$$p_{2n,2n} = \mathbb{P}[T_{2n} = 2n] = \mathbb{P}[S_1 \geq 0, \ldots, S_{2n} \geq 0] = \mathbb{P}[S_1 \leq 0, \ldots, S_{2n} \leq 0] = \mathbb{P}[M_n = 0],$$

wobei wir die Symmetrie der Irrfahrt ausgenutzt haben und $M_n = \max\{S_0, S_1, \ldots, S_n\}$. Mit (14.3.1) erhalten wir

$$p_{2n,2n} = 1 - \mathbb{P}[M_{2n} \geq 1] = 1 - 2\mathbb{P}[S_{2n} \geq 1] + \mathbb{P}[S_{2n} = 1] = \mathbb{P}[S_{2n} = 0] = u_{2n},$$

denn $\mathbb{P}[S_{2n} = 1] = 0$ und $2\mathbb{P}[S_{2n} \geq 1] = \mathbb{P}[|S_{2n}| \geq 1]$. Somit gilt die Formel $p_{2n,2n} = u_{2n}u_0$. Analog erhält man im Fall $k = 0$

$$p_{0,2n} = \mathbb{P}[T_{2n} = 0] = \mathbb{P}[S_1 \leq 0, \ldots, S_{2n} \leq 0] = \mathbb{P}[M_n = 0] = u_{2n} = u_0 u_{2n}.$$

Schritt 1. Sei nun $k = 1, \ldots, n-1$. Wir wollen die Wahrscheinlichkeit bestimmen, dass $T_{2n} = 2k$. Wir betrachten die erste Rückkehr zum Ursprung und bezeichnen den entsprechenden Zeitpunkt mit $2r$. Es gibt zwei Fälle:

Fall A. Die erste Rückkehr findet zum Zeitpunkt $2r$ statt, nachdem sich die Irrfahrt auf der positiven Halbachse aufgehalten hat. Die Wahrscheinlichkeit dieses Ereignisses ist $\frac{1}{2} f_{2r}$. Also hat die Irrfahrt bereits $2r$ Zeiteinheiten auf der positiven Halbachse verbracht und muss noch $2k - 2r$ Zeiteinheiten dort verbringen. Dabei stehen ihr insgesamt $2n - 2r$ Zeiteinheiten zu Verfügung. Es muss also $2r \leq 2k$ und $2n - 2r \geq 2k - 2r$ gelten. Wir erhalten somit den Term

$$\frac{1}{2} \sum_{r=1}^{k} f_{2r} p_{2k-2r,2n-2r}.$$

Fall B. Die erste Rückkehr findet zum Zeitpunkt $2r$ statt, nachdem sich die Irrfahrt auf der negativen Halbachse aufgehalten hat. Die Wahrscheinlichkeit dieses Ereignisses ist $\frac{1}{2} f_{2r}$. Danach muss die Irrfahrt noch $2k$ Zeiteinheiten auf der nichtnegativen Halbachse verbringen.

Dafür stehen ihr $2n-2r$ Zeiteinheiten zur Verfügung. Es muss also $2k \le 2n-2r$ gelten. Wir erhalten somit den Term

$$\frac{1}{2}\sum_{r=1}^{n-k} f_{2r} p_{2k,2n-2r}.$$

Beide Fälle zusammengenommen ergeben die Identität

$$p_{2k,2n} = \frac{1}{2}\sum_{r=1}^{k} f_{2r} p_{2k-2r,2n-2r} + \frac{1}{2}\sum_{r=1}^{n-k} f_{2r} p_{2k,2n-2r}. \tag{14.4.1}$$

Nun beenden wir den Beweis mittels Induktion.
Induktionsbasis: Für die Fälle $n=1, k=0$ und $n=1, k=1$ lässt sich die Aussage leicht durch Abzählen der Pfade überprüfen:

$$p_{0,2} = p_{2,2} = \frac{1}{2}, \quad u_0 = 1, \quad u_2 = \frac{1}{2}.$$

Induktionsannahme: Es gelte

$$p_{2k,2m} = u_{2k} \cdot u_{2m-2k} \quad \text{für alle } m = 1, \dots, n-1 \text{ und } k = 1, \dots, m-1. \tag{14.4.2}$$

Nun ist zu zeigen, dass diese Formel auch für $m = n$ und beliebiges $k = 1, \dots, n-1$ gilt. Mit (14.4.1) und (14.4.2) erhalten wir

$$\begin{aligned}
p_{2k,2n} &= \frac{1}{2}\sum_{r=1}^{k} f_{2r} \cdot u_{2k-2r} \cdot u_{2n-2k} + \frac{1}{2}\sum_{r=1}^{n-k} f_{2r} \cdot u_{2k} \cdot u_{2n-2k-2r} \\
&= \frac{1}{2} u_{2n-2k} \sum_{r=1}^{k} f_{2r} \cdot u_{2k-2r} + \frac{1}{2} u_{2k} \sum_{r=1}^{n-k} f_{2r} \cdot u_{2n-2k-2r} \\
&= u_{2k} \cdot u_{2n-2k}.
\end{aligned}$$

Dabei haben wir im letzten Schritt die Identitäten

$$\sum_{r=1}^{k} f_{2r} u_{2k-2r} = u_{2k} \text{ und } \sum_{r=1}^{n-k} f_{2r} u_{2n-2k-2r} = u_{2n-2k},$$

siehe (14.2.3), benutzt. Somit gilt $p_{2k,2n} = u_{2k} \cdot u_{2n-2k}$ und die Induktion ist beendet. □

Satz 14.4.2 (Arcussinus–Gesetz). *Die Zufallsvariable $\frac{T_{2n}}{2n}$ konvergiert für $n \to \infty$ in Verteilung gegen eine Zufallsvariable $T$ mit der Dichte*

$$f_T(y) = \frac{1}{\pi\sqrt{y(1-y)}}, \quad y \in [0,1].$$

Bemerkung 14.4.3. Hierbei ist $\frac{T_{2n}}{2n}$ ist der Anteil der Zeit, die die Irrfahrt auf der positiven Halbachse verbringt. Der Satz heißt Arcussinus–Gesetz (und die Zufallsvariable $T$ heißt Arcussinus-verteilt), da die Verteilungsfunktion von $T$ die folgende Gestalt hat:

$$F_T(y) = \frac{2}{\pi} \arcsin\sqrt{y}, \quad y \in [0,1].$$

Die Dichte $f_T(y)$ erreicht ihr Minimum bei $t = 1/2$. Für $t \to 0$ und $t \to 1$ wird sie unendlich. Es ist demnach unwahrscheinlich, dass die Irrfahrt auf der positiven und negativen Halbachse ungefähr gleich viel Zeit verbringt. Viel wahrscheinlicher ist es, dass die Irrfahrt entweder fast die ganze Zeit auf der positiven, oder fast die ganze Zeit auf der negativen Halbachse verbringt.

Beweis von Satz 14.4.2. Hier soll nur eine Beweisidee vorgestellt werden. Wir haben in Satz 14.4.1 und Satz 14.2.8 gezeigt, dass

$$\mathbb{P}[T_{2n} = 2k] = p_{2k,2n} = u_{2k} \cdot u_{2n-2k}, \quad u_{2k} = \mathbb{P}[S_{2k} = 0] \sim \frac{1}{\sqrt{\pi k}}, \quad k \to \infty.$$

Es seien $a$ und $b$ zwei Zahlen mit $0 < a < b < 1$. Wir erhalten, dass

$$\begin{aligned}
\mathbb{P}\left[a < \frac{T_{2n}}{2n} \le b\right] &= \sum_{k=[na]+1}^{[nb]} u_{2k} \cdot u_{2n-2k} \\
&\sim \frac{1}{n} \sum_{k=[na]+1}^{[nb]} \frac{1}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}} \qquad \text{(Riemann-Summe)} \\
&\underset{n\to\infty}{\longrightarrow} \int_a^b \frac{dy}{\pi\sqrt{y(1-y)}}.
\end{aligned}$$

Dabei ist die Begründung des Übergangs von der ersten zur zweiten Zeile eine Übung. $\square$

Außer Satz 14.4.2 gibt es mindestens zwei weitere Arcussinus–Gesetze, die wir ohne Beweis angeben. In den beiden Sätzen ist $S_0, S_1, S_2, \ldots$ eine Irrfahrt mit $p = 1/2$ und $T$ ist eine Arcussinus-verteilte Zufallsvariable.

Satz 14.4.4 (Arcussinus-Gesetz für die letzte Nullstelle). *Es sei* $K_{2n} = \max\{k : S_k = 0, 0 \le k \le 2n\}$ *die letzte Nullstelle der Irrfahrt im Zeitintervall* $[0, 2n]$. *Dann gilt*

$$\frac{K_{2n}}{2n} \underset{n\to\infty}{\overset{d}{\longrightarrow}} T.$$

Satz 14.4.5 (Arcussinus–Gesetz für die Position des Maximums). *Es sei* $R_{2n} = \max\{k : S_k = M_k, 0 \le k \le 2n\}$ *die Position des letzten Maximums der Irrfahrt im Zeitintervall* $[0, 2n]$. *Dann gilt*

$$\frac{R_{2n}}{2n} \underset{n\to\infty}{\overset{d}{\longrightarrow}} T.$$

## 14.5. Gesetz vom iterierten Logarithmus

Seien $X_1, X_2, \ldots$ unabhängige identisch verteilte Zufallsvariablen mit $\mathbb{E}X_k = 0$ und $\operatorname{Var} X_k = 1$. Definiere

$$S_n = X_1 + \ldots + X_n, \quad n \in \mathbb{N}.$$

Wir haben folgende Aussagen über die Wachstumsgeschwindigkeit der Folge $S_1, S_2, \dots$ bewiesen.

1. Das starke Gesetz der großen Zahlen behauptet, dass

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{f.s.} 0.$$

Die Folge $S_n$ wächst also wesentlich langsamer als $n$.

2. Der zentrale Grenzwertsatz behauptet, dass

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} N, \quad N \sim N(0,1).$$

Der Wert $S_n$ ist also mit großer Wahrscheinlichkeit ungefähr von der Größenordnung $\sqrt{n}$.

3. In Satz 14.3.1 haben wir gezeigt, dass

$$\frac{\max\{S_0, S_1, \dots, S_n\}}{\sqrt{n}} \xrightarrow[n\to\infty]{d} |N|, \quad N \sim N(0,1).$$

Wir haben nur den Fall $\mathbb{P}[X_k = \pm 1] = 1/2$ betrachtet, die Aussage gilt aber allgemein. Das Maximum von $S_0, S_1, \dots, S_n$ ist also mit großer Wahrscheinlichkeit von der Größenordnung $\sqrt{n}$.

Welche ist nun die richtige Geschwindigkeit, mit der die Folge $S_1, S_2, S_3, \dots$ wächst? Diese Frage beantwortet der folgende Satz.

Satz 14.5.1 (Gesetz vom iterierten Logarithmus). *Es seien $X_1, X_2, \dots$ unabhängige identisch verteilte Zufallsvariablen mit $\mathbb{E}X_k = 0$ und $\operatorname{Var} X_k = 1$. Definiere $S_n = X_1 + \ldots + X_n$, $n \in \mathbb{N}$. Dann gilt*

$$\mathbb{P}\left[\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = 1\right] = \mathbb{P}\left[\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\log(\log n)}} = -1\right] = 1.$$

Ohne Beweis.

Beispiel 14.5.2. Aus dem Gesetz vom iterierten Logarithmus folgt, dass für alle $\varepsilon > 0$ mit Wahrscheinlichkeit 1 gilt:

$$\lim_{n\to\infty} \frac{S_n}{n^{\frac{1}{2}+\varepsilon}} = 0 \text{ jedoch } \limsup_{n\to\infty} \frac{S_n}{n^{\frac{1}{2}-\varepsilon}} = +\infty.$$

Die Folge $S_n$ wächst also langsamer als $n^{\frac{1}{2}+\varepsilon}$, aber nicht langsamer als $n^{\frac{1}{2}-\varepsilon}$.