KAPITEL 8

Rekorde

Seien $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen mit stetiger Verteilungsfunktion F. Wir setzen $M_n = \max\{X_1, \ldots, X_n\}$.

Definition 8.0.1. Wir sagen, dass zum Zeitpunkt $n \in \mathbb{N}$ ein **Rekord** aufgestellt wird, wenn der Wert X_n größer als alle vorherigen Werte $X_1, \ldots, X_{n-1}$ ist.

Der zum Zeitpunkunkt $n = 1$ beobachtete Wert gilt per Definition immer als ein Rekord. Wir definieren deshalb die **Rekord–Indikatoren** $\xi_1, \xi_2, \ldots$ durch $\xi_1 = 1$ und

$$\xi_n = \mathbb{1}_{X_n > M_{n-1}}, \qquad n = 2, 3, \ldots.$$

Somit ist ξ_n die Indikatorvariable des Ereignisses, dass zum Zeitpunkt n ein neuer Rekord aufgestellt wird.

8.1. Satz von Rényi

Der folgende Satz beschreibt die gemeinsame Verteilung der Zufallsvariablen $\xi_1, \xi_2, \ldots$.

Satz 8.1.1 (Rényi, 1962, und Dwass, 1960)**.** Es gilt $\mathbb{P}[\xi_n = 1] = \frac{1}{n}$ für alle $n \in \mathbb{N}$. Außerdem sind die Zufallsvariablen $\xi_1, \ldots, \xi_n, M_n$ unabhängig.

BEWEIS. SCHRITT 1. Wir zeigen, dass $\mathbb{P}[\xi_n = 1] = \frac{1}{n}$. Aus der Stetigkeit der Verteilungsfunktion F folgt, dass $\mathbb{P}[X_i = X_j] = 0$ für $i \neq j$. Somit sind alle Werte $X_1, X_2, \ldots$ unterschiedlich mit Wahrscheinlichkeit 1. Es gilt

$$1 = \mathbb{P}[M_n = X_1] + \mathbb{P}[M_n = X_2] + \ldots + \mathbb{P}[M_n = X_n] = n\mathbb{P}[M_n = X_n] = n\mathbb{P}[\xi_n = 1],$$

wobei die erste Gleichheit wegen der Disjunktheit der Ereignisse gilt und die zweite Gleichheit aus Symmetriegründen Bestand hat. (Jede der Beobachtungen $X_1, \ldots, X_n$ ist mit gleicher Wahrscheinlichkeit $\frac{1}{n}$ das Maximum). Es folgt, dass $\mathbb{P}[\xi_n = 1] = \frac{1}{n}$.

SCHRITT 2. Wir zeigen die Unabhängigkeit von $\xi_1, \ldots, \xi_n, M_n$. Seien dazu $1 \leq \alpha(1) < \alpha(2) < \ldots < \alpha(s) \leq n$ und $x \in \mathbb{R}$ beliebig. Es reicht zu zeigen, dass

$$\mathbb{P}[\xi_{\alpha(1)} = 1, \ldots, \xi_{\alpha(s)} = 1, M_n < x] = \mathbb{P}[\xi_{\alpha(1)} = 1] \cdot \ldots \cdot \mathbb{P}[\xi_{\alpha(s)} = 1] \cdot \mathbb{P}[M_n < x].$$

Es sei zuerst $s = 1$. Schreibe $M_{k,l} = \max\{X_k, \ldots, X_l\}$ mit $k \leq l$. Um die Notation zu vereinfachen, schreiben wir α anstelle von $\alpha(1)$. Es gilt

$$\mathbb{P}[\xi_\alpha = 1, M_n < x] = \mathbb{P}[M_{\alpha-1} < X_\alpha < x, M_{\alpha+1,n} < x] = \mathbb{P}[M_{\alpha-1} < X_\alpha < x](F(x))^{n-\alpha},$$

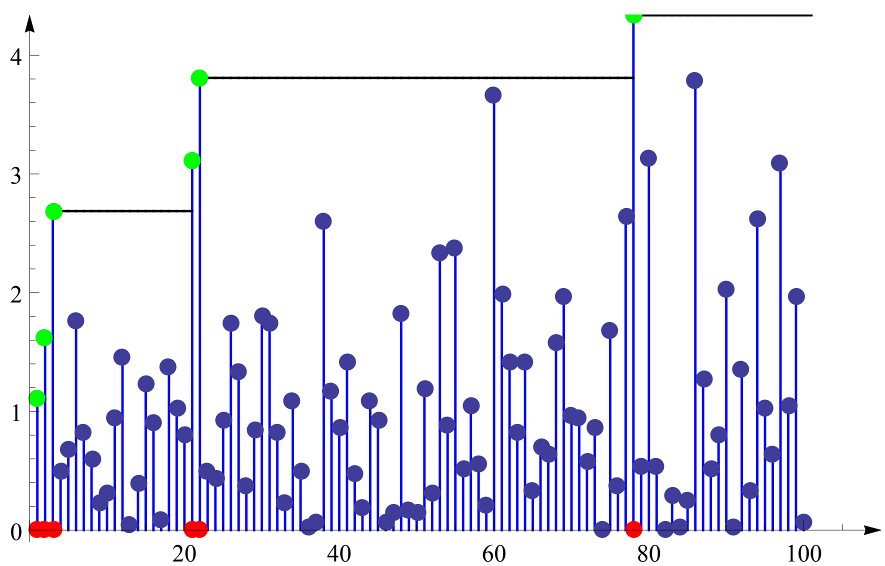

ABBILDUNG 1. Rekorde. Rote Punkte: Rekordzeiten $L(n)$. Grüne Punkte: Rekordwerte $X(n)$.

wobei in der letzten Gleichheit die Unabhängigkeit benutzt wurde. Indem man nun auf $X_\alpha = u \in (-\infty, x)$ bedingt und die Formel der totalen Wahrscheinlichkeit unter Berücksichtigung von $\mathbb{P}[M_{\alpha-1} < u] = (F(u))^{\alpha-1}$ verwendet, kann man das wie folgt umschreiben:

$$\mathbb{P}[\xi_\alpha = 1, M_n < x] = (F(x))^{n-\alpha} \cdot \int_{-\infty}^{x} (F(u))^{\alpha-1} \mathrm{d}F(u).$$

Setzt man nun $w = F(u)$, so erhält man, dass

$$\mathbb{P}[\xi_\alpha = 1, M_n < x] = (F(x))^{n-\alpha} \cdot \int_0^{F(x)} w^{\alpha-1} \mathrm{d}w = \frac{(F(x))^n}{\alpha} = \mathbb{P}[M_n < x] \cdot \mathbb{P}[\xi_\alpha = 1],$$

was die Aussage im Fall $s = 1$ beweist.

Sei nun $s \in \mathbb{N}$ beliebig. Es gilt

$$\begin{aligned}\mathbb{P}[\xi_{\alpha(1)} = 1, \ldots, \xi_{\alpha(s)} = 1, M_n < x] = \mathbb{P}[&M_{\alpha(1)-1} < X_{\alpha(1)} < x, M_{\alpha(1),\ldots,\alpha(2)-1} < X_{\alpha(2)} < x, \\ &\ldots, M_{\alpha(s-1),\ldots,\alpha(s)-1} < X_{\alpha(s)} < x, M_{\alpha(s)+1,n} < x].\end{aligned}$$

Indem man nun auf $X_{\alpha(1)} = u_1, \ldots, X_{\alpha(s)} = u_s$ bedingt, kann man den obigen Ausdruck ähnlich wie im Fall $s = 1$ schreiben als

$$(F(x))^{n-\alpha(s)} \int_{u_1 < u_2 < \ldots < u_s < x} (F(u_1))^{\alpha(1)-1} \ldots (F(u_s))^{\alpha(s)-\alpha(s-1)-1} \mathrm{d}F(u_1) \ldots \mathrm{d}F(u_s).$$

Setzt man $F(u_1) = w_1, \ldots, F(u_s) = w_s$, so erhält man

$$(F(x))^{n-\alpha(s)} \int_{0 < w_1 < w_2 < \ldots < w_s < F(x)} w_1^{\alpha(1)-1} \ldots w_s^{\alpha(s)-\alpha(s-1)-1} \mathrm{d}w_1 \ldots \mathrm{d}w_s.$$

Als Übungsaufgabe bleibt zu zeigen, dass sich obiges Integral zu Folgendem errechnen lässt:

$$\frac{F^n(x)}{\alpha(1) \cdot \ldots \cdot \alpha(s)} = \mathbb{P}[M_n < x] \cdot \mathbb{P}[\xi_{\alpha(1)=1}] \cdot \ldots \cdot \mathbb{P}[\xi_{\alpha(s)=1}].$$

(Man kann z.B. Induktion nach s verwenden). $\square$

Bemerkung 8.1.2. Nach dem Satz von Rényi hängt die gemeinsame Verteilung von $\xi_1, \xi_2, \ldots$ nicht von der Verteilungsfunktion F ab.

8.2. Anzahl der Rekorde

Es sei $N(n)$ die **Anzahl der Rekorde** im Intervall $1, \ldots, n$, d.h.

$$N(n) = \sum_{k=1}^{n} \xi_k.$$

Nach dem Satz von Rényi gilt

$$\mathbb{E}N(n) = \sum_{k=1}^{n} \mathbb{E}\xi_k = \sum_{k=1}^{n} \frac{1}{k} = \log n + \gamma + o(1),$$

$$\operatorname{Var} N(n) = \sum_{k=1}^{n} \operatorname{Var} \xi_k = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k^2}\right) = \log n + \gamma - \frac{\pi^2}{6} + o(1),$$

für $n \to \infty$. Dabei ist γ die **Euler–Mascheroni–Konstante**:

$$\gamma = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \log n\right) = 0,57721\ldots.$$

Es werden also unter den ersten n Beobachtungen lediglich ungefähr $\log n$ Rekorde (was sehr wenig ist!) erwartet.

Im nächsten Satz werden wir die komplette Verteilung von $N(n)$ mit Hilfe von Stirling–Zahlen erster Art beschreiben.

Definition 8.2.1. Die **Stirling–Zahlen erster Art** sind definiert als Koeffizienten in der Formel

$$x(x+1)\ldots(x+n-1) = \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} x^k.$$

Der nächste Satz beschreibt die Verteilung der Anzahl der Rekorde $N(n)$.

Satz 8.2.2. Für die Verteilung der Anzahl der Rekorde gilt

$$\mathbb{P}[N(n) = k] = \frac{1}{n!} \begin{bmatrix} n \\ k \end{bmatrix}, \qquad k = 1, \ldots, n.$$

Bemerkung 8.2.3. Setzt man $x = 1$ in die Definition der Stirling–Zahlen ein, so erhält man

$$\sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} = n!.$$

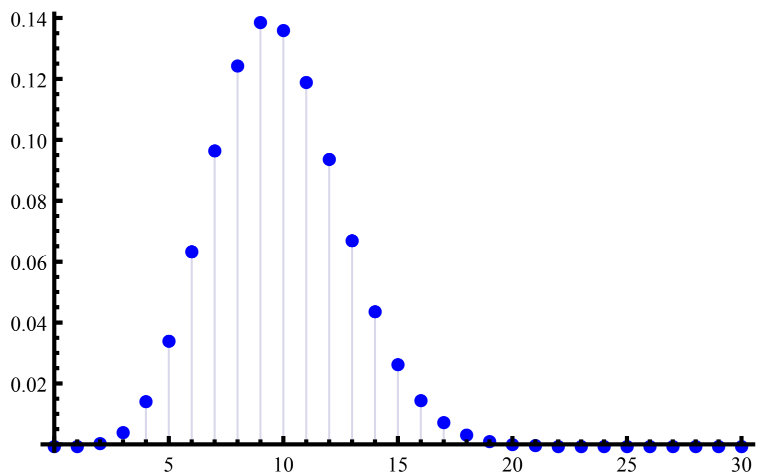

Abbildung 2. Zähldichte der Zufallsvariable $N(n)$ für $n = 10000$.

Also summieren sich die Wahrscheinlichkeiten zu 1.

Beweis. Die Zufallsvariablen $\xi_1, \ldots, \xi_n$ sind nach dem Satz von Rényi unabhängig und Bernoulli–verteilt mit

$$\mathbb{P}[\xi_k = 1] = \frac{1}{k}, \quad \mathbb{P}[\xi_k = 0] = 1 - \frac{1}{k}.$$

Die erzeugende Funktion einer Zufallsvariable Z mit Werten in $\{0, 1, \ldots\}$ ist definiert durch

$$g_Z(t) = \mathbb{E}[t^Z] = \sum_{k=0}^{\infty} \mathbb{P}[Z = k] t^k, \qquad |t| < 1.$$

Im Folgenden werden wir die erzeugende Funktion von $N(n)$ angeben. Die erzeugende Funktion von ξ_k ist

$$g_{\xi_k}(t) = \left(1 - \frac{1}{k}\right) + \frac{t}{k}.$$

Es gilt $N(n) = \xi_1 + \ldots + \xi_n$ (mit unabhängigen Summanden) und deshalb ist

$$g_{N(n)}(t) = g_{\xi_1}(t) \ldots g_{\xi_n}(t) = \prod_{k=1}^{n} \frac{k - 1 + t}{k} = \frac{t(t+1)\ldots(t+n-1)}{n!} = \frac{1}{n!} \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} t^k,$$

wobei im letzten Schritt die Definition der Stirling–Zahlen verwendet wurde. Auf der anderen Seite gilt definitionsgemäß

$$g_{N(n)}(t) = \sum_{k=1}^{n} t^k \mathbb{P}[N(n) = k].$$

Durch Vergleich der Koeffizienten erhalten wir die gewünschte Formel. □

Aufgabe 8.2.4. Zeigen Sie, dass die Stirling–Zahlen erster Art die folgende Rekursionsformel erfüllen:

$$\begin{bmatrix} n+1 \\ k \end{bmatrix} = n \begin{bmatrix} n \\ k \end{bmatrix} + \begin{bmatrix} n \\ k-1 \end{bmatrix}.$$

Zum Vergleich: Für Binomialkoeffizienten gilt eine ähnliche Formel ohne den Faktor n:

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}.$$

Aufgabe 8.2.5. Beweisen Sie für die Stirling–Zahl erster Art die Formel

$$\frac{1}{n!}\begin{bmatrix} n+1 \\ k+1 \end{bmatrix} = \sum \frac{1}{i_1 \dots i_k},$$

wobei über alle ganzzahligen $1 \leq i_1 < i_2 < \dots < i_k \leq n$ summiert wird.

8.3. Rekordzeiten

Wir definieren nun die **Rekordzeiten** $L(1) < L(2) < \dots$ durch: $L(1) = 1$, $L(2) = \min\{j > 1 : \xi_j = 1\}$ und allgemein

$$L(n+1) = \min\{j > L(n) : \xi_j = 1\}, \qquad n = 2, 3, \dots.$$

Somit ist $L(n)$ der Zeitpunkt, zu dem der n-te Rekord aufgestellt wird. Im nächsten Satz beschreiben wir die gemeinsame Verteilung des Vektors $(L(1), \dots, L(n))$.

Satz 8.3.1. Für beliebige natürliche Zahlen $1 = j(1) < j(2) < \dots < j(n)$ gilt

$$\mathbb{P}[L(1) = j(1), L(2) = j(2), \dots, L(n) = j(n)] = \frac{1}{j(n)(j(2)-1)\dots(j(n)-1)}.$$

Beweis. Es gilt:

$$\mathbb{P}[L(1) = j(1), \dots, L(n) = j(n)] = \mathbb{P}[\xi_2 = \xi_3 = \dots = \xi_{j(2)-1} = 0,$$
$$\xi_{j(2)} = 1, \xi_{j(2)+1} = \dots = \xi_{j(3)-1} = 0, \xi_{j(3)} = 1, \dots, \xi_{j(n)} = 1].$$

Wegen der sich aus Satz 8.1.1 ergebenden Unabhängigkeit kann man dies in folgenden Ausdruck umschreiben:

$$\mathbb{P}[\xi_2 = 0] \cdot \ldots \cdot \mathbb{P}[\xi_{j(n)} = 0] \cdot \frac{\mathbb{P}[\xi_{j(2)} = 1]}{\mathbb{P}[\xi_{j(2)} = 0]} \cdot \ldots \cdot \frac{\mathbb{P}[\xi_{j(n)} = 1]}{\mathbb{P}[\xi_{j(n)} = 0]},$$

was sich ebenfalls, wegen des Satzes von Rényi, wie folgt darstellen lässt:

$$\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right) \dots \left(1 - \frac{1}{j(n)}\right) \frac{1/j(2)}{1 - 1/j(2)} \dots \frac{1/j(n)}{1 - 1/j(n)}.$$

Durch geschicktes Umformen lässt sich das wie folgt darstellen:

$$\frac{2-1}{2}\frac{3-1}{3} \dots \frac{j(n)-1}{j(n)} \frac{1}{j(2)-1} \dots \frac{1}{j(n)-1} = \frac{1}{j(n)} \frac{1}{j(2)-1} \dots \frac{1}{j(n)-1},$$

wobei sich die Gleichheit ergibt, da die ersten $j(n)-1$ Faktoren ein Teleskopprodukt bilden. □

Bemerkung 8.3.2. Die gemeinsame Verteilung der Rekordzeiten $L(1), L(2), \dots$ ist (abgesehen von der Stetigkeitsannahme an die Verteilungsfunktion F) unabhängig von der Verteilung der Zufallsvariablen $X_1, X_2, \dots$.

Bemerkung 8.3.3. Die Verteilung von $L(2)$ sieht somit folgendermaßen aus:

$$\mathbb{P}[L(2) = j] = \frac{1}{j(j-1)}, \quad j = 2, 3, \ldots.$$

Insbesondere gilt $\mathbb{E}L(2) = \sum_{j=1}^{\infty} \frac{1}{j-1} = +\infty$. Die *mittlere* Wartezeit auf den zweiten Rekord ist somit unendlich. (Was erstaunlich ist!) Auf der anderen Seite ist die Wartezeit auf den zweiten Rekord *fast sicher* endlich.

Satz 8.3.4. Für die Verteilung der Rekordzeit $L(n)$ gilt

$$P[L(n) = k] = \frac{1}{k!} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}, \qquad k = n, n+1, \ldots.$$

Beweis. Übung: Benutzen Sie Satz 8.2.2. □

Im nächsten Satz werden wir zeigen, dass die Rekordzeiten $L(1), L(2), \ldots$ eine **Markov–Kette** bilden.

Satz 8.3.5. Die Folge $L(1), L(2), \ldots$ ist eine Markov–Kette mit Anfangszustand $L(1) = 1$ und Übergangswahrscheinlichkeiten

$$p_{ij} = \frac{i}{j(j-1)}$$

für $i = 1, 2, \ldots$ und $j = i+1, i+2, \ldots$.

Beweis. Zu zeigen ist, dass für alle $1 = j(1) < j(2) < \ldots < j(n)$ gilt

$$\mathbb{P}[L(1) = j(1), L(2) = j(2), \ldots, L(n) = j(n)] = p_{j(1)j(2)} p_{j(2)j(3)} \ldots p_{j(n-1)j(n)}.$$

Mit der Formel aus Satz 8.3.1 gilt

$$\begin{aligned}
&\mathbb{P}[L(1) = j(1), L(2) = j(2), \ldots, L(n) = j(n)] \\
&= \frac{1}{j(n)(j(2)-1)\ldots(j(n)-1)} \\
&= \frac{j(1)}{j(2)(j(2)-1)} \cdot \frac{j(2)}{j(3)(j(3)-1)} \cdot \ldots \cdot \frac{j(n-1)}{j(n)(j(n)-1)},
\end{aligned}$$

was die gewünschte Formel ergibt. □

Angenommen, die ersten n Rekordzeiten sind bekannt: $L(1) = 1$, $L(2) = i(2), \ldots, L(n) = i(n)$. Wo liegt nun die nächste Rekordzeit $L(n+1)$? Wegen der Markov–Eigenschaft der Folge $L(1), L(2), \ldots$ stellt es sich heraus, dass man für die Bestimmung von $L(n+1)$ lediglich den Wert $L(n) = i(n)$ benötigt, die Werte von $L(1), \ldots, L(n-1)$ sind hingegen irrelevant.

Satz 8.3.6. Für alle $1 = i(1) < i(2) < \ldots < i(n) = i < j$ gilt die Markov–Eigenschaft der Rekordzeiten:

$\mathbb{P}[L(n+1) = j | L(n) = i] = \mathbb{P}[L(n+1) = j | L(n) = i, L(n-1) = i(n-1), \ldots, L(2) = i(2)]$.

Außerdem gilt:

$$\mathbb{P}[L(n+1) = j | L(n) = i] = \frac{i}{j(j-1)}.$$

Beweis. Folgt aus Satz 8.3.5. □

Der obige Satz zeigt, wie man die Folge der Rekordzeiten am Rechner simulieren kann, ohne dafür die Variablen $X_1, X_2, \ldots$ erzeugen zu müssen. Man startet mit $L(1) = 1$ und geht induktiv vor. Sind die Werte $L(1), \ldots, L(n)$ mit $L(n) = i$ bekannt, so erzeugt man eine Zufallsvariable auf $\{i+1, i+2, \ldots\}$, indem man den Wert j mit Wahrscheinlichkeit $\frac{j}{i(i-1)}$ auswählt. Dieser Wert ist dann der Wert von $L(n+1)$. Danach wiederholt man das Ganze.

Der nächste Satz gibt einen viel einfacheren Algorithmus zur Simulation der Rekordzeiten. Definiere die Gauß–Klammer durch

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}, \qquad x \in \mathbb{R}.$$

Satz 8.3.7 (Williams, 1973)**.** Seien $U_1, U_2, \ldots$ unabhängig und gleichverteilt auf dem Intervall $[0, 1]$. Definiere $R(1) = 1$ und $R(n+1) = \lfloor \frac{R(n)}{U_n} \rfloor + 1$ für $n \in \mathbb{N}$. Dann gilt die Gleichheit der Verteilungen:

$$(L(1), L(2), \ldots, L(n)) \stackrel{d}{=} (R(1), R(2), \ldots, R(n)).$$

Beweis. Wegen der Markov–Eigenschaft reicht es zu zeigen, dass für alle $i \in \mathbb{N}$ und $j \in \{i+1, i+2, \ldots\}$ gilt

$$\mathbb{P}[L(n+1) = j | L(n) = i] = \mathbb{P}[R(n+1) = j | R(n) = i].$$

Die Wahrscheinlichkeit auf der linken Seite ist gleich $\frac{i}{j(j-1)}$ nach Satz 8.3.6. Wir berechnen die Wahrscheinlichkeit auf der rechten Seite. Es ist

$$\mathbb{P}[R(n+1) = j | R(n) = i] = \mathbb{P}\left[\left\lfloor \frac{R(n)}{U_n} \right\rfloor + 1 = j | R(n) = i\right] = \mathbb{P}\left[\left\lfloor \frac{i}{U_n} \right\rfloor + 1 = j | R(n) = i\right].$$

Die Zufallsvariable $R(n)$ hängt nur von $U_1, \ldots, U_{n-1}$ ab. Die Ereignisse $\{\lfloor \frac{i}{U_n} \rfloor + 1 = j\}$ und $\{R(n) = i\}$ sind also unabhängig und somit vereinfacht sich das Ganze zu folgendem Ausdruck:

$$\mathbb{P}[R(n+1) = j | R(n) = i] = \mathbb{P}\left[\left\lfloor \frac{i}{U_n} \right\rfloor = j - 1\right] = \mathbb{P}\left[\frac{i}{U_n} \in [j-1, j)\right] = \frac{i}{j(j-1)},$$

da U_n gleichverteilt auf $[0, 1]$ ist. □

Satz 8.3.8 (Tata, 1969)**.** Sei $x > 1$ eine natürliche Zahl, dann gilt:
$$\mathbb{P}\left[\frac{L(n+1)}{L(n)} > x\right] = \frac{1}{x}.$$
Sei $x > 1$ eine beliebige reelle Zahl, dann gilt:
$$\lim_{n\to\infty} \mathbb{P}\left[\frac{L(n+1)}{L(n)} > x\right] = \frac{1}{x}.$$

Bemerkung 8.3.9. Man sieht hier noch einmal, dass Rekorde mit $n \to \infty$ immer seltener auftreten. Zum Beispiel ist mit Wahrscheinlichkeit $1/2$ der Abstand zwischen der $(n+1)$-ten und der n-ten Rekordzeit größer als die n-te Rekordzeit selbst. Mit Wahrscheinlichkeit $1/3$ ist der Abstand zwischen der $(n+1)$-ten und der n-ten Rekordzeit mindestens doppelt so groß wie die n-te Rekordzeit selbst, usw. Die Tatsache, dass Rekorde immer seltener auftreten ist ziemlich natürlich: die Rekordwerte steigen nämlich mit der Zeit und es wird immer schwieriger neue Rekorde aufzustellen.

Beweis. Sei $x > 1$. Dann gilt wegen des Gesetzes der totalen Wahrscheinlichkeit:
$$\begin{aligned}\mathbb{P}[L(n+1) > xL(n)] &= \sum_{i=n}^{\infty} \mathbb{P}[L(n+1) > xi | L(n) = i] \cdot \mathbb{P}[L(n) = i] \\ &= \sum_{i=n}^{\infty} \mathbb{P}[L(n+1) > [xi] | L(n) = i] \cdot \mathbb{P}[L(n) = i],\end{aligned}$$
denn $L(n+1)$ ist ganzzahlig und somit ist $L(n+1) > xi$ zu $L(n+1) > [xi]$ äquivalent. Im Beweis von Satz 8.3.6 wurde die Wahrscheinlichkeit $\mathbb{P}[L(n+1) > xi|L(n) = i]$ bereits berechnet. Wir wollen das Ergebnis hier verwenden. So lässt sich obiger Ausdruck zu folgendem vereinfachen:
$$\mathbb{P}[L(n+1) > xL(n)] = \sum_{i=n}^{\infty} \frac{i}{[xi]} \mathbb{P}[L(n) = i]. \tag{8.3.1}$$
Sei zuerst $x \in \mathbb{N}$. Man sieht:
$$\sum_{i=n}^{\infty} \frac{i}{[xi]} \mathbb{P}[L(n) = i] = \frac{1}{x} \sum_{i=n}^{\infty} \mathbb{P}[L(n) = i] = \frac{1}{x}.$$
Sei nun $x > 1$ beliebig reell. Aus der Definition der Gauß–Klammer folgt, dass $[xi] \le xi < [xi] + 1$. Durch leichte Umformungen folgt:
$$\frac{1}{x} \le \frac{i}{[xi]} < \frac{1}{x} + \frac{1}{[xi]x}.$$
Deshalb kann man (8.3.1) wie folgt nach unten abschätzen:
$$\sum_{i=n}^{\infty} \frac{i}{[xi]} \mathbb{P}[L(n) = i] \ge \frac{1}{x} \sum_{i=n}^{\infty} \mathbb{P}[L(n) = i] = \frac{1}{x}.$$

Außerdem kann man daher (8.3.1) wie folgt nach oben abschätzen:

$$\sum_{i=n}^{\infty} \frac{i}{[xi]} \mathbb{P}[L(n) = i] < \sum_{i=n}^{\infty} \left(\frac{1}{x} + \frac{1}{[xi]x} \right) \mathbb{P}[L(n) = i] \leq \frac{1}{x} + \frac{1}{x} \frac{1}{[xn]},$$

da $[xi] \geq [xn]$ für $i \geq n$. Insgesamt folgt mit dem "Sandwich–Prinzip" die Behauptung. □

Korollar 8.3.10. Sei U gleichverteilt auf $[0, 1]$, dann gilt

$$\frac{L(n+1)}{L(n)} \xrightarrow[n\to\infty]{d} \frac{1}{U}.$$

Beweis. Die Zufallsvariable $\frac{1}{U}$ ist Pareto–verteilt mit Tailfunktion $\frac{1}{x}$, $x > 1$. Die Behauptung folgt nun aus Satz 8.3.8. □

Bemerkung 8.3.11. Shorrock, 1972, hat gezeigt, dass für beliebiges $k \in \mathbb{N}$ und für unabhängige, auf $[0, 1]$ gleichverteilte Zufallsvariablen $U_1, \ldots, U_k$, gilt

$$\left(\frac{L(n+1)}{L(n)}, \frac{L(n+2)}{L(n+1)}, \ldots, \frac{L(n+k)}{L(n+k-1)} \right) \xrightarrow[n\to\infty]{d} \left(\frac{1}{U_1}, \frac{1}{U_2}, \ldots, \frac{1}{U_k} \right).$$

8.4. Zentrale Grenzwertsätze

Nun zeigen wir, dass für großes n die Anzahl der Rekorde $N(n)$ approximativ normalverteilt ist, siehe Abbildung 2. Mit $\mathcal{N}(0, 1)$ bezeichnen wir die Starndardnormalverteilung mit Verteilungsfunktion

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-\frac{1}{2}z^2} dz.$$

Satz 8.4.1 (Zentraler Grenzwertsatz für die Anzahl der Rekorde)**.** Es gilt

$$\frac{N(n) - \log n}{\sqrt{\log n}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, 1) \text{ für } n \to \infty.$$

Der Beweis basiert auf einer Verallgemeinerung des zentralen Grenzwertsatzes, die wir ohne Beweis angegeben.

Satz 8.4.2 (Zentraler Grenzwertsatz von Ljapunow)**.** Für jedes $n \in \mathbb{N}$ seien $Z_{n1}, \ldots, Z_{nn}$ unabhängige Zufallsvariablen mit $\mathbb{E}Z_{nk} = 0$, $\sigma_{nk}^2 := \operatorname{Var} Z_{nk} \in (0, \infty)$ für $k = 1, \ldots, n$

und $\sum_{k=1}^n \sigma_{nk}^2 = 1$. Außerdem gelte die folgende *Ljapunow–Bedingung*: Für ein $\delta > 0$

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}|Z_{nk}|^{2+\delta} \to 0.$$

Dann gilt:

$$Z_{n1} + \ldots + Z_{nn} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1).$$

Bemerkung 8.4.3. Die identische Verteiltheit der Zufallsvariablen wird im zentralen Grenzwertsatz von Ljapunow nicht vorausgesetzt.

Beweis von Satz 8.4.1. Es ist $N(n) = \xi_1 + \ldots + \xi_n$, wobei die $\xi_1, \ldots, \xi_n$ nach dem Satz von Rényi unabhängig aber nicht identisch verteilt sind. Setze

$$Z_{nk} = \frac{\xi_k - \frac{1}{k}}{\sigma_n}, \qquad k = 1, \ldots, n,$$

mit

$$\sigma_n^2 = \operatorname{Var} N(n) = \sum_{k=1}^n \operatorname{Var} \xi_k = \sum_{k=1}^n \frac{1}{k} - \sum_{k=1}^n \frac{1}{k^2} \sim \log n, \qquad n \to \infty,$$

denn $\sum_{k=1}^n \frac{1}{k} \sim \log n$ und $\sum_{k=1}^\infty \frac{1}{k^2} < \infty$. Definitionsgemäß gilt

$$\mathbb{E} Z_{nk} = 0, \qquad \sum_{k=1}^n \mathbb{E} Z_{nk}^2 = 1.$$

Wir zeigen, dass die Ljapunow–Bedingung mit $\delta = 1$ gilt. Die Zufallsvariable $\frac{\xi_k - 1/k}{\sigma_n}$ nimmt nur zwei Werte an, und zwar $\frac{1/k}{\sigma_n}$ mit Wahrscheinlichkeit $1 - \frac{1}{k}$ und $\frac{1-1/k}{\sigma_n}$ mit Wahrscheinlichkeit $\frac{1}{k}$. Es folgt

$$\sum_{k=1}^n \mathbb{E}\left|\frac{\xi_k - \frac{1}{k}}{\sigma_n}\right|^3 \le \sum_{k=1}^n \left(\frac{1}{k\sigma_n}\right)^3 \cdot 1 + \left(\frac{1}{\sigma_n}\right)^3 \cdot \frac{1}{k} = \frac{1}{\sigma_n^3} \sum_{k=1}^n \left(\frac{1}{k^3} + \frac{1}{k}\right) \to 0 \text{ für } n \to \infty,$$

da $\sigma_n^3 \sim (\log n)^{3/2}$ und $\sum_{k=1}^n \left(\frac{1}{k^3} + \frac{1}{k}\right) \sim \log n$ für $n \to \infty$. Es sind also alle Voraussetzungen für den zentralen Grenzwertsatz von Ljapunow gegeben und damit folgt:

$$\frac{N(n) - \sum_{k=1}^n \frac{1}{k}}{\sigma_n} = \sum_{k=1}^n Z_{nk} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1) \text{ für } n \to \infty.$$

Unter Beachtung des bereits gezeigten Zusammenhangs $\sigma_n^2 \sim \log n$ führt das zur behaupteten Grenzaussage:

$$\frac{N(n) - \log n}{\sqrt{\log n}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1) \text{ für } n \to \infty.$$

(Man kann z.B. das Lemma von Chintschin verwenden). □

Aufgabe 8.4.4. Zeigen Sie, dass für die Anzahl der Rekorde auch das Gesetz der großen Zahlen gilt:

$$\lim_{n\to\infty} \frac{N(n)}{\log n} = 1 \qquad \text{f.s.}$$

Nun benutzen wir den obigen Satz um auch einen zentralen Grenzwertsatz für die Rekordzeiten herzuleiten.

Satz 8.4.5 (Zentraler Grenzwertsatz für Rekordzeiten)**.** Es gilt

$$\frac{\log L(n) - n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1).$$

Beweis. Sei $x \in \mathbb{R}$ fest. Mit $n(x) = \mathrm{e}^{n+x\sqrt{n}}$ gilt

$$\mathbb{P}\left[\frac{\log L(n) - n}{\sqrt{n}} \leq x\right] = \mathbb{P}[L(n) \leq n(x)] = \mathbb{P}[N(n(x)) \geq n].$$

Zur Vereinfachung sei $n(x) \in \mathbb{Z}$, dabei verliert der Beweis nicht an Allgemeingültigkeit, da für $n(x) \notin \mathbb{Z}$ einfach $[n(x)]$ betrachtet werden kann. Nach Satz 8.4.1 gilt

$$\frac{N(n(x)) - \log n(x)}{\sqrt{\log n(x)}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1).$$

Es folgt, dass

$$\lim_{n\to\infty} \mathbb{P}[N(n(x)) \geq n] = \lim_{n\to\infty} \mathbb{P}\left[\frac{N(n(x)) - \log n(x)}{\sqrt{\log n(x)}} \geq \frac{n-(n+x\sqrt{n})}{\sqrt{n+x\sqrt{n}}}\right] = 1-\Phi(-x) = \Phi(x),$$

da $\lim_{n\to\infty} \frac{n-(n+x\sqrt{n})}{\sqrt{n+x\sqrt{n}}} = -x$. Dabei bezeichnet $\Phi(x)$ die Verteilungsfunktion von $\mathcal{N}(0,1)$ und die Behauptung folgt. $\square$

Aufgabe 8.4.6. Zeigen Sie, dass für die Rekordzeiten auch das Gesetz der großen Zahlen gilt:

$$\lim_{n\to\infty} \frac{\log L(n)}{n} = 1 \qquad \text{f.s.}$$

Aufgabe 8.4.7. Ist es richtig, dass

$$\lim_{n\to\infty} \frac{L(n)}{\mathrm{e}^n} = 1 \qquad \text{f.s.?}$$

8.5. Rekordwerte

Definition 8.5.1. Die **Rekordwerte** sind definiert als

$$X(n) = M_{L(n)} = X_{L(n)}.$$

Anders als bei Rekordindikatoren oder Rekordzeiten, hängt die Verteilung der Rekordwerte von der Verteilungsfunktion der Zufallsvariablen X_i ab. Im Spezialfall der exponentialverteilten Zufallsvariablen besitzen die Rekordwerte eine schöne Darstellung.

Satz 8.5.2 (Tata, 1969)**.** Seien $X_1, X_2, \dots$ unabhängig und exponentialverteilt mit Parameter 1. Dann gilt:

$$(X(1), X(2), \dots, X(n)) \stackrel{d}{=} (\nu_1, \nu_1 + \nu_2, \dots, \nu_1 + \dots + \nu_n),$$

wobei $\nu_1, \nu_2, \dots$ unabhängig unabhängig und exponentialverteilt mit Parameter 1 sind.

Bemerkung 8.5.3. Mit anderen Worten, die Folge $X(1), X(2), \dots$ ist ein Poisson–Prozess mit Intensität 1.

Beweisidee. Der Beweis basiert auf der Gedächtnislosigkeit der Exponentialverteilung. Ist nämlich $X \sim \text{Exp}(1)$, so ist für jedes $t > 0$ die bedingte Verteilung von $X - t$ gegeben, dass $X > t$, ebenfalls eine Exponentialverteilung mit Parameter 1.

Nach Voraussetzung ist $X(1) = X_1 \sim \text{Exp}(1)$. Wir halten nun $X_1 = a_1$ fest und warten auf den zweiten Rekordwert $X(2)$. Der Exzess $X(2) - a_1$ hat die gleiche Verteilung wie $X - a_1$ gegeben, dass $X > a_1$, also die Exponentialverteilung mit Parameter 1. Nun halten wir $X(2) = a_2$ fest und warten auf den dritten Rekordwert $X(3)$. Der Exzess $X(3) - a_2$ hat die gleiche Verteilung, wie $X - a_2$ gegeben, dass $X > a_2$, also wieder die Exponentialverteilung mit Parameter 1, usw.

Somit sind die Zuwächse $X(1), X(2) - X(1), X(3) - X(2), \dots$ standard exponentialverteilt und unabhängig. $\square$

Korollar 8.5.4. Seien $X_1, X_2, \dots$ unabhängig und exponentialverteilt mit Parameter 1. Dann gilt

$$\frac{X(n) - n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1) \text{ für } n \to \infty.$$

Beweis. Aus Satz 8.5.2 wissen wir, dass $X(n) \stackrel{d}{=} \nu_1 + \dots + \nu_n$. Dabei ist $\mathbb{E}\nu_i = \operatorname{Var} \nu_i = 1$. Der zentrale Grenzwertsatz ergibt, dass

$$\frac{\nu_1 + \dots + \nu_n - n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1) \text{ für } n \to \infty.$$

Daraus folgt die Behauptung. $\square$

Bemerkung 8.5.5. In Satz **??** haben wir nachgewiesen, dass im Fall der standard exponentialverteilten Zufallsvariablen, die Zufallsvariable $M_n - \log n$ gegen die Gumbel–Verteilung Λ konvergiert. Es gilt also

$$M_n - \log n \xrightarrow[n\to\infty]{d} \Lambda \quad \text{und} \quad \frac{M_{L(n)} - n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1).$$

Die Grenzverteilung wird also offenbar durch den Umstand, dass $L(n)$ zufällig ist, völlig verändert.

Aufgabe 8.5.6. Seien $X_1, X_2, \ldots$ unabhängige identisch verteilte Zufallsvariablen mit stetiger Verteilungsfunktion F. Es sei $X(n)$ der n-te Rekordwert.

(1) Zeigen Sie, dass
$$\mathbb{P}[X(n) < x] = Q_n(F(x)), \quad x \in \mathbb{R},$$
wobei $Q_n(s) = \mathbb{E}s^{L(n)}$ die erzeugende Funktion von $L(n)$ sei.

(2) Zeigen Sie, dass
$$Q_n(s) = \frac{1}{(n-1)!} \int_0^{-\log(1-s)} t^{n-1} \mathrm{e}^{-t} \mathrm{d}t.$$

Hinweis: Zu Teil (2): Teil (1) gilt auch für exponentialverteilte Zufallsvariablen.