

6 Rekorde

Seien X_1, X_2, \dots unabhängige und identisch verteilte Zufallsvariablen mit stetiger Verteilungsfunktion F . Wir setzen $M_n = \max\{X_1, \dots, X_n\}$. Seien ξ_1, ξ_2, \dots Zufallsvariablen mit $\xi_1 = 1$ und

$$\xi_n = 1_{X_n > M_{n-1}}, \quad n = 2, 3, \dots$$

Das Ereignis $\{\xi_n = 1\}$ tritt genau dann ein, wenn der Wert X_n größer als alle vorherigen Werte X_1, \dots, X_{n-1} ist. Somit ist ξ_n die Indikatorvariable des Ereignisses, dass zum Zeitpunkt n ein neuer Rekord aufgestellt wird.

Theorem 6.1 (Rényi). *Es gilt $\mathbb{P}[\xi_n = 1] = \frac{1}{n}$. Außerdem sind die Zufallsvariablen ξ_1, \dots, ξ_n, M_n unabhängig.*

Beweis. Aus der Stetigkeit von F folgt, dass $\mathbb{P}[X_i = X_j] = 0$ für $i \neq j$. Es gilt

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

wobei die erste Gleichheit wegen der Disjunktheit der Ereignisse gilt und die zweite Gleichheit aus Symmetriegründen Bestand hat. Es folgt, dass $\mathbb{P}[\xi_n = 1] = \frac{1}{n}$.

Wir zeigen die Unabhängigkeit. Seien nun $1 \leq \alpha(1) < \alpha(2) < \dots < \alpha(s) \leq n$ und $x \in \mathbb{R}$. Wir beweisen, dass

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

Es sei zuerst $s = 1$. Schreibe $M_{k,l} = \max\{X_k, \dots, X_l\}$ mit $k \leq l$. Wir zeigen, dass

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

Dies kann man wie folgt mit Hilfe der Identitäten

$$\mathbb{P}[X_{\alpha(1)} = u] = F(du), \quad \mathbb{P}[M_{\alpha(1)-1} < u] = (F(u))^{\alpha(1)-1}$$

umschreiben:

$$\mathbb{P}[M_{\alpha(1)-1} < X_{\alpha(1)} < x](F(x))^{n-\alpha(1)} = (F(x))^{n-\alpha(1)} \cdot \int_{-\infty}^x F(du)(F(u))^{\alpha(1)-1}.$$

Das Integral lässt sich einfach zu folgendem Ausdruck integrieren:

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

was die Aussage im Fall $s = 1$ beweist. Sei nun $s \in \mathbb{N}$. Es gilt

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

Dies lässt sich ähnlich wie im Fall $s = 1$ schreiben als:

$$(F(x))^{n-\alpha(s)} \int_{-\infty}^x F(du_1)(F(u_1))^{\alpha(1)-1} \dots \int_{u_{s-1}}^x F(du_s)(F(u_s))^{\alpha(s)-\alpha(s-1)-1}.$$

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

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

□

Bemerkung 6.2. Es sei $N(n) = \sum_{k=1}^n \xi_k$ die Anzahl der Rekorde im Intervall $1, \dots, n$. Nach dem Satz von Rényi gilt

$$\mathbb{E}N(n) = \sum_{k=1}^n \mathbb{E}\xi_k = \sum_{k=1}^n \frac{1}{k} \sim \log n, \quad n \rightarrow \infty.$$

Es werden also relativ wenige Rekorde erwartet.

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

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

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

Theorem 6.3. Für $1 = j(1) < j(2) < \dots < j(n)$ 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 6.1 ergebenden Unabhängigkeit kann man dies in folgenden Ausdruck umschreiben:

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

was sich ebenfalls wegen dem Satz von Rényi als folgender Ausdruck 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. \square

Bemerkung 6.4. Die Verteilung der Rekordzeiten $L(1), L(2), \dots$ ist (abgesehen von der Stetigkeitsannahme an die Verteilungsfunktion F) also unabhängig von der Verteilung der Zufallsvariablen X_1, X_2, \dots

Bemerkung 6.5. Wir können nun die Verteilung von $L(n)$ ausrechnen. Es gilt

$$\mathbb{P}[L(n) = k] = \sum_{1=j(1) < j(2) < \dots < j(n)=k} \mathbb{P}[L(1) = j(1), \dots, L(n) = j(n)].$$

Wegen Satz 6.3 erhalten wir

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

Wir werden später eine explizite Formel für die Verteilung von $L(n)$ herleiten.

Im nächsten Satz werden wir zeigen, dass die Folge der Rekordzeiten $L(1), L(2), \dots$ eine sogenannte Markov-Kette bildet. Angenommen, die ersten n Rekordzeiten sind bekannt: $L(1) = 1, L(2) = i(2), \dots, L(n) = i(n)$. Wo liegt nun die nächste Rekordzeit $L(n+1)$? Es stellt sich heraus, dass man für die Bestimmung von $L(n+1)$ lediglich den Wert $L(n) = i(n)$ benötigt, die Werte von $L(1), \dots, L(n-1)$ sind irrelevant. Diese Eigenschaft wird als die Markov-Eigenschaft bezeichnet.

Theorem 6.6. Für alle $1 = i(1) < i(2) < \dots < i(n) = i < j$ gilt die Markov-Eigenschaft der Rekordzeiten:

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

Außerdem gilt:

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

Beweis. Wir berechnen nur die Wahrscheinlichkeit $\mathbb{P}[L(n+1) = j | L(n) = i]$, für die gilt:

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

was folgendem Ausdruck entspricht:

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

Nach Satz 6.3 lässt sich dies umformen zu:

$$\frac{\sum_{1=i(1) < j(2) < \dots < i(n)=i < j=i(n+1)} \frac{1}{j \cdot (i(2)-1) \cdot \dots \cdot (i(n-1)-1) \cdot (i-1) \cdot (j-1)}}{\sum_{1=i(1) < j(2) < \dots < i(n)=i < j=i(n+1)} \frac{1}{i \cdot (i(2)-1) \cdot \dots \cdot (i(n-1)-1) \cdot (i-1)}} = \frac{i}{j(j-1)}.$$

□

Der obige Satz zeigt, wie man die Folge der Rekordzeiten am Rechner simulieren kann, ohne dafür die Variablen X_1, X_2, \dots erzeugen zu müssen. Man startet mit $L(1) = 1$ und geht induktiv vor. Sind die Werte $L(1), \dots, L(n)$ mit $L(n) = i$ bekannt, so erzeugt man eine Zufallsvariable auf $\{i+1, i+2, \dots\}$ 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 liefert eine einfache Darstellung der Rekordzeiten. Definiere die Gauß-Klammer durch

$$[x] = \max\{n \in \mathbb{Z} : n \leq x\}, \quad x \in \mathbb{R}.$$

Theorem 6.7. Seien U_1, U_2, \dots unabhängig gleichverteilt auf dem Intervall $[0, 1]$. Definiere $R(1) = 1$ und $R(n+1) = \lceil \frac{R(n)}{U_n} \rceil + 1$ für $n \in \mathbb{N}$. Dann gilt die Gleichheit der Verteilungen:

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

Beweis. Wegen der Markov-Eigenschaft reicht es zu zeigen, dass

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

Es ist

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

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

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

Da es sich bei U_n um eine auf $[0, 1]$ gleichverteilte Zufallsvariable handelt ergibt sich:

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

Der letzte Term ist gemäß Satz 6.6 gleich der Wahrscheinlichkeit $\mathbb{P}[L(n+1) = j | L(n) = i]$, was zu zeigen war. \square

Wir haben schon gezeigt, dass für $N(n) = \sum_{k=1}^n \xi_k$ (die Anzahl der Rekorde im Intervall $\{1, \dots, n\}$) gilt $\mathbb{E}N(n) = \sum_{k=1}^n \frac{1}{k}$. Im nächsten Satz berechnen wir die komplette Verteilung von $N(n)$. Dabei stoßen wir auf die Stirling-Zahlen S_n^k , die durch die Gleichung

$$x(x-1)\dots(x-n+1) = \sum_{k=0}^n S_n^k x^k$$

definiert sind.

Theorem 6.8. *Für die Verteilung der Anzahl der Rekorde gilt*

$$\mathbb{P}[N(n) = k] = (-1)^{n+k} \frac{S_n^k}{n!} = \frac{|S_n^k|}{n!}, \quad k = 1, \dots, n.$$

Bemerkung 6.9. Setzt man $x = -1$ in die Definition der Stirling-Zahlen ein, so erhält man $\sum_{k=1}^n (-1)^{n+k} S_n^k = n!$. Also summieren sich die Wahrscheinlichkeiten zu 1.

Beweis. Die Zufallsvariablen ξ_1, \dots, ξ_n sind nach dem Satz von Rényi unabhängig mit $\mathbb{P}[\xi_k = 1] = \frac{1}{k}$ bzw. $\mathbb{P}[\xi_k = 0] = 1 - \frac{1}{k}$.

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

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

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

$$g_{\xi_m}(t) = \left(1 - \frac{1}{m}\right) + t \cdot \frac{1}{m}.$$

Es gilt $N(n) = \xi_1 + \dots + \xi_n$ und deshalb ist:

$$g_{N(n)}(t) = g_{\xi_1}(t) \dots g_{\xi_n}(t) = \prod_{k=1}^n \frac{k-1+t}{k} = (-1)^n \prod_{k=1}^n \frac{-t-(k-1)}{k}.$$

Mit Hilfe der Definition der Stirlingzahlen lässt sich dies zu folgendem Ausdruck umformen:

$$g_{N(n)}(t) = \frac{(-1)^n}{n!} \sum_{k=1}^n S_n^k (-t)^k = \sum_{k=1}^n t^k S_n^k \frac{(-1)^{n+k}}{n!}.$$

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

$$\mathbb{P}[N(n) = k] = (-1)^{n+k} \frac{S_n^k}{n!} = \frac{|S_n^k|}{n!},$$

wobei die letzte Gleichheit aus der Nichtnegativität der Wahrscheinlichkeiten folgt. \square

Aufgabe 6.10. Zeigen Sie mit dem obigen Satz, dass für die Verteilung der Rekordzeiten gilt

$$P[L(n) = k] = \frac{|S_{n-1}^{k-1}|}{k!}, \quad k = n, n+1, \dots$$

Mit $\mathcal{N}(0, 1)$ bezeichnen wir die Standardnormalverteilung mit Verteilungsfunktion

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

Theorem 6.11 (Zentraler Grenzwertsatz für die Anzahl der Rekorde). *Es gilt*

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

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

Theorem 6.12 (Zentraler Grenzwertsatz von Ljapunow). *Für alle natürlichen Zahlen n seien Z_{n1}, \dots, Z_{nn} unabhängige Zufallsvariablen mit $\mathbb{E}Z_{nk} = 0$, $\sigma_{nk}^2 := \text{Var}Z_{nk} \in (0, \infty)$ für $k = 1, \dots, n$ und $\sum_{k=1}^n \sigma_{nk}^2 = 1$. Außerdem gelte $\sum_{k=1}^n \mathbb{E}|Z_{nk}|^{2+\delta} \rightarrow 0$ für ein $\delta > 0$ und $n \rightarrow \infty$. Dann gilt:*

$$Z_{n1} + \dots + Z_{nn} \xrightarrow{d} \mathcal{N}(0, 1) \text{ für } n \rightarrow \infty.$$

Bemerkung 6.13. Die identische Verteiltheit der Zufallsvariablen wird im zentralen Grenzwertsatz von Ljapunow nicht benötigt und daher nicht vorausgesetzt.

Beweis von Satz 6.11. Es ist $N(n) = \xi_1 + \dots + \xi_n$, wobei die ξ_k nach dem Satz von Rényi unabhängig aber nicht identisch verteilt sind. Setze

$$Z_{nk} = \frac{\xi_k - \frac{1}{k}}{\sigma_n}, \quad k = 1, \dots, n,$$

mit

$$\sigma_n^2 = \text{Var}(N(n)) = \sum_{k=1}^n \text{Var}(\xi_k) = \sum_{k=1}^n \frac{1}{k} - \sum_{k=1}^n \frac{1}{k^2} \sim \log n, \quad n \rightarrow \infty,$$

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

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

Wir zeigen, dass die Ljapunow-Bedingung mit $\delta = 3$ 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

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

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 $n \rightarrow \infty$. Es sind also alle Voraussetzungen für den zentralen Grenzwertsatz von Ljapunow gegeben und damit folgt:

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

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

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

(Man kann z.B. das Khintchine-Lemma verwenden). □

Mithilfe des obigen Satzes leiten wir einen zentralen Grenzwertsatz für die Rekordzeiten her.

Theorem 6.14 (Zentraler Grenzwertsatz für Rekordzeiten). *Es gilt*

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

Beweis. Sei $x \in \mathbb{R}$ fest. Mit $n(x) = e^{n+x\sqrt{n}}$ 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 6.11 gilt

$$\frac{N(n(x)) - \log n(x)}{\sqrt{\log n(x)}} \rightarrow \mathcal{N}(0, 1) \text{ für } n \rightarrow \infty.$$

Es folgt, dass

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

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

Aufgabe 6.15. Zeigen Sie, dass für die Anzahl der Rekorde auch das Gesetz der großen Zahlen gilt:

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

Theorem 6.16. 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:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{L(n+1)}{L(n)} > x \right] = \frac{1}{x}.$$

Das heißt, $\frac{L(n+1)}{L(n)}$ konvergiert in Verteilung gegen eine Pareto-verteilte Zufallsvariable mit Verteilungsfunktion $1 - \frac{1}{x}$, $x > 1$.

Bemerkung 6.17. Man sieht hier noch einmal, dass Rekorde mit $n \rightarrow \infty$ 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 $L(n+1) > [xi]$ equivalent. Im Beweis zum Satz 6.6 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:

$$\mathbb{P}[L(n+1) > xL(n)] = \sum_{i=n}^{\infty} \frac{i}{[xi]} \mathbb{P}[L(n) = i]. \quad (6.1)$$

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 $[xi] \leq xi < [xi]+1$. Durch leichte Umformungen folgt:

$$\frac{1}{x} \leq \frac{i}{[xi]} < \frac{1}{x} + \frac{1}{[xi]x}.$$

Deshalb kann man (6.1) wie folgt nach unten abschätzen:

$$\sum_{i=n}^{\infty} \frac{i}{[xi]} \mathbb{P}[L(n) = i] \geq \frac{1}{x} \sum_{i=n}^{\infty} \mathbb{P}[L(n) = i] = \frac{1}{x}.$$

Außerdem kann man daher (6.1) 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. \square

Aufgabe 6.18. Zeigen Sie, dass für die Rekordzeiten auch das Gesetz der großen Zahlen gilt:

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

Aufgabe 6.19. Ist es richtig, dass

$$\lim_{n \rightarrow \infty} \frac{N(n)}{e^n} = 1 \quad \text{f.s.?}$$

Zum Schluss betrachten wir noch die Rekordwerte. Diese sind definiert als

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

Im Spezialfall der exponentialverteilten Zufallsvariablen besitzen die Rekordwerte eine schöne Darstellung.

Theorem 6.20. Seien X_1, X_2, \dots unabhängig und mit Parameter 1 exponentialverteilte Zufallsvariablen. Dann gilt:

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

wobei ν_1, ν_2, \dots unabhängig unabhängig und exponentialverteilt mit Parameter 1 sind.

Bemerkung 6.21. Mit anderen Worten, die Folge $X(1), X(2), \dots$ ist ein Poisson-Prozess mit Intensität 1.

Beweisidee. Der Beweis basiert auf der Vergessenseigenschaft 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.

Anschaulich kann man die Verteilung der Rekordwerte $X(1), X(2), \dots$ wie folgt erklären: Die Zuwächse im Bild 6.1 sind immer standard exponentialverteilt und unabhängig voneinander.

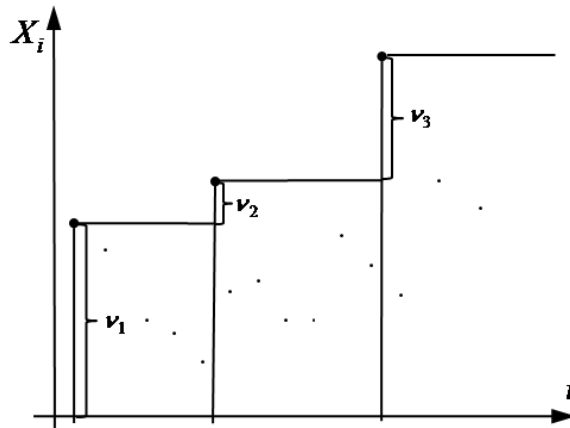


Abbildung 6.1: Darstellung der Zuwächse $X(i) - X(i - 1)$

□

Korollar 6.22. *Unter Voraussetzungen des obigen Satzes gilt:*

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

Beweis. Aus Satz 6.20 wissen wir $X(n) \stackrel{d}{=} \nu_1 + \dots + \nu_n$. Dabei ist $\mathbb{E}\nu_i = \text{Var } \nu_i = 1$. Der zentrale Grenzwertsatz ergibt, dass

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

Daraus folgt die Behauptung. □

Bemerkung 6.23. Im Satz ?? haben wir nachgewiesen, dass für standard exponentialverteilte Zufallsvariablen $M_n - \log n$ gegen die Gumbel-Verteilung konvergiert. Unter passenden Normierungen gilt also $M_n \rightarrow G(t)$ und gleichzeitig $M_{L(n)} \rightarrow \mathcal{N}(0, 1)$. Die Grenzverteilung wird also offenbar durch den Umstand, dass $L(n)$ zufällig ist völlig verändert.