SEMINARARBEIT

The Inverse Gamma Distribution and Benford's Law

ausgeführt am

Institut für
Finanz- und Versicherungsmathematik
TU Wien

unter der Anleitung von

Assoc. Prof. Dipl.-Ing. Dr.techn. Stefan Gerhold

durch

Lisa Retzer
Matrikelnummer: 11802445

Wien, am 28.2.2021

Inhaltsverzeichnis

1 Einleitung **1**

2 Das Gesetz von Benford **2**
2.1 Ursprung 2
2.2 Das Gesetz von Benford 2

3 Die Inverse Gammaverteilung **6**

4 Die Annäherung der Inversen Gamma Verteilung an Benford's Law **7**
4.1 Reihenentwicklung für F'_N 7
4.2 Fehlerabschätzung 13
4.3 Intepretation und Graphiken 18

5 Appendix **21**

Abbildungsverzeichnis **22**

Literaturverzeichnis **23**

1 Einleitung

Das Gesetz von Benford und die Inverse Gammaverteilung sind unter anderem zwei wichtige Bestandteile in der Finanz-Mathematik. Während das Gesetz von Benford sehr bekannt in der Fraud-Detection ist, findet sich die Inverse Gamma Verteilung beispielsweise bei Stresstests wieder.

In dieser Seminararbeit wird untersucht, unter welche Bedingungen an die Parameter der Inversen Gammaverteilung eine Annäherung an das Gesetz von Benford erfolgt.

Diese Analyse besteht aus zwei Hauptteilen:

- Zum einen wird eine Reihenentwicklung für die Ableitung einer Zufallsvariable aufgestellt, die durch die Inversen Gammaverteilung konstruiert wird.
- Und andererseits wird für diese Reihenentwicklung eine Fehlerschranke berechnet.

Im letzten Abschnitt dieser Arbeit werden die Resultate in mehreren Grafiken visualisiert. Hier wird dann ersichtlich, dass die Inverse Gammaverteilung je näher am Gesetz von Benford ist, desto kleiner der verwendete Parameter α ist. Außerdem wird man sehen, dass für kleine α der Parameter β kaum ein Einfluss auf die Abweichung hat. Auch bei einem größeren α bleibt die Abweichung ungefähr in derselben Größenordnung, wenn β variiert wird.

2 Das Gesetz von Benford

2.1 Ursprung

Der erste Artikel, dessen Inhalt mit dem Gesetz von Benford in Verbndung steht, wurde 1881 von dem kanadischen Astronom und Mathematiker Simon Newcomb im *American Journal of Mathematics* veröffentlicht. In seinem Beitrag, names *Note on the Frequency of Use of the Different Digits in Natural Numbers*, berichtet er, es sei ihm die Ungleichverteilung der führenden Ziffern in Logarithmus Tafeln aufgefallen, woraufhin er sich mit der Verteilung der Ziffern in einer Zahl beschäftigt hat.

Als Hauptaussage führt Newcomb an, die Verteilung von Zahlen sei durch die Gleichverteilung der Mantissen ihres Logarimuses gegeben. Diese Definition für das Gesetz von Benford wird in Satz 2.5 nocheinmal aufgegriffen.

Newcomb hat zusätzlich eine Tabelle mit der Verteilung der Ziffern veröffentlicht. Die angegebenen Wahrscheinlichkeiten sind auf 4 Nachkommastellen angegeben und stimmen für die führende Ziffer mit jenen Resultaten von Frank Benford, die erst 50 Jahre später gemacht werden, überein.

Auch Frank Benford, ein US-Amerikanischer Physiker und Elektroingenier, sind Logarithmus Tafeln ins Auge gefallen. Bei Benford war es aber eher der Grad der Abnutzung der ersten Seiten, der für ihn besonders auffällig war. Infolgedessen hat er 20 Datensätze auf die führende Ziffer hin untersucht, die von der Länge von Flüssen und Physikalische Größen über Todesfallraten bis hin zu mathematischen Folgen gereicht haben.

In seiner Veröffentlichung *The Law of Anomalous Numbers* in den *Proceedings of the American Philosophical Society* von 1938 finden sich seine Ergebnisse der Analyse der Datensätze. Einige weichen deutlich von der Benford Verteilung ab, andere sie sehr gut, trotzdem ist der Durchschnitt aller Datenwert wieder sehr nahe am Gesetz von Benford. [M, Seiten 4ff.]

2.2 Das Gesetz von Benford

Bevor das Gesetz von Benford formal definiert wird, wird die Notation für die Mantisse geklärt.

Bemerkung 2.1. Sei $N \geq 2$ eine Natürliche Zahl, dann hat jede postive reelle Zahl $x \in \mathbb{R}$ eine Darstellung der Form $x = S_N(x) \cdot N^{i(x)}$, wobei $S_N(x) \in [1, N)$ die *Mantisse* von x zur Basis N und $i(x)$ eine Ganze Zahl ist. [D, Seite 1]

Wie auch im Artikel von Newcomb enthält das Gesetz von Benford eine Aussage über die vorkommenden Mantissen innerhalb eines natürlich entstehenden Datensatzes. Es postuliert, dass für jedes $1 \leq s < N$ der relative Anteil des Datensatzes, dessen Mantisse höchstens s ist, gleich $\log_N(s)$ beträgt.

Definition 2.2. (Gesetz von Benford, [D, Seite 1f]) Sei X eine Zufallsvariable mit Werten in $(0, \infty)$. Dann befolgt X das Gesetz von Benford zur Basis N, falls für alle $s \in [1, N)$ gilt:

$$\mathbb{P}[\, S_N(X) \leq s \,] = \log_N(s).$$

Insbesondere gilt:

$$\mathbb{P}[\, \text{die erste Ziffer von } X \text{ ist } z \,] = \log_N\left(\frac{z+1}{z}\right)$$

Im Beispiel von Bemerkung 2.1, wäre X die Verteilung der Zahlen im Datensatz, N die Basis und S_N die Verteilung der Mantissen zur Basis N.

Für $N = 10$ können die Zahlen aus einem Datensatz besonders schnell in die Gleitkommadarstellung mit Zehnerpotenzen geschrieben werden. Hier ist die erste Ziffer der Mantisse auch die führende Ziffer der ursprünglichen Zahl. Beispielweise kann $125,56$ auch als $1,2345 \cdot 10^2$ geschrieben werden. $123,56$ würde somit in die Klasse der Zahlen fallen, die mit 1 beginnen bzw. deren führende Ziffer die 1 ist.

Die nachstehende Abbildung visualisiert die Verteilung der führenden Ziffer zur Basis 10. Je höher die Zahl, desto seltener kommt sie als führende Ziffer vor. Die Wahrscheinlichkeit nach Benford, dass eine Zahl mit der 1 beginnt, liegt bei $30,1$ % während nur $4,58$ % der Zahlen mit einer 9 beginnen müssten.

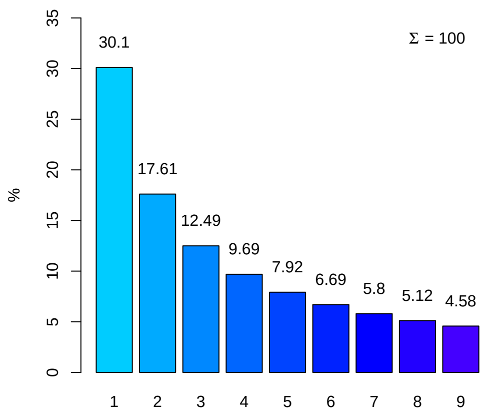


Abbildung 2.1: Verteilung der führenden Ziffer zur Basis 10
Quelle: eigene Darstellung

Ab hier wird in den nachstehenden Kaptieln mit einer anderen Definition des Gesetzes von Benford gearbeitet. Dafür wird die Gleichverteilung benötigt.

Bemerkung 2.3 (Stetige Gleichverteilung, [H])**.** Sei Z eine Zufallsvariable, dann nennt man Z (stetig) gleichverteilt auf $[a,b]$, falls ihre Dichtefunktion die folgende Form hat:

$$f_Z(x) = \frac{1}{b-a} \cdot \mathbb{1}_{[a,b]},$$

wobei $\mathbb{1}_{[a,b]}$ die Indikatorfunktion über dem Intervall $[a,b]$ bezeichnet.

Bemerkung 2.4. Sei Z eine (stetig) gleichverteilte Zufallsvariable auf $[0,1]$, dann ordnet sie jedem Teilintervall $[a,b] \subseteq [0,1] \subset \mathbb{R}$ seine Länge zu:

$$\mathbb{P}\Big[\, Z \in [a,b] \,\Big] = b - a$$

[D, Seite 2]

Im folgenden Satz wird die Notation: $Z \mod 1$ verwendet, wobei Z eine Zufallsvariable ist. Unter der Rechenoperation "Modulo" versteht man den Divisionsrest. Die Zufallsvariable $Z \mod 1$ gibt also den Divisionrest bzgl. 1 wieder. Das sind genau die Nachkommastellen. Beispielsweise gilt:

$$\begin{aligned} 3,5 \mod 1 &= 0,5 \\ 1,7 \mod 1 &= 0,7 \end{aligned}$$

Nun kann eine äquivalente Definition für das Gesetz von Benford aufgestellt werden:

Satz 2.5. *Eine Zufallsvariable X folgt dem Gesetz von Benford, falls die Zufallsvariable $Y := \log_N(X) \mod 1$ gleichverteilt auf $[0,1]$ ist. [D, Seite 2]*

Beweis. Sei $Y := \log_N(X) \mod 1$ gleichverteilt auf $[a,b]$, dann gilt:

$$\begin{aligned} Y &= \log_N(X) \mod 1 \\ &= \log_N(S_N(X) \cdot N^{i(X)}) \mod 1 && (2.1) \\ &= \log_N(S_N(X)) + \log_N(N^{i(X)}) \mod 1 && (2.2) \\ &= \log_N(S_N(X)) + i(X) \mod 1 && (2.3) \\ &= \log_N(S_N(X)) && (2.4) \end{aligned}$$

Da der Logorithmus nur für positive Werte definiert ist, darf X nur positive Werte annehmen. Nach Bemerkung 2.1 kann jede positive reelle Zahl als Produkt aus Mantisse $(S_N(X))$ und Potenz $(N^{i(X)})$ zu einer Basis N geschrieben werden, wordurch (2.1) gilt.

(2.2) ist auf die Logarithmus-Rechenregel $\log(a \cdot b) = \log(a) + \log(b)$ zurückzuführen.

Außerdem gilt wegen folgenden Logarithmus-Rechenregeln die Gleichung in (2.3):

$$\log_N(N^{i(X)}) = i(X) \cdot \log_N(N) = i(X) \cdot 1 = i(X)$$

Laut Bemerkung 2.1 ist $i(X)$ eine ganze Zahl, und da die Addition mit einer ganze Zahl die Modulo-Rechnung bzgl. 1 nicht beeinflusst, kann diese weggelassen werden.

Die Gleichverteilung von $Y := \log_N(X) \mod 1$ ist daher äquivalent zur Gleichverteilung von $\log_N(S_N(X))$ und nach Bemerkung 2.4 gilt für alle $1 \leq p \leq N$:

$$\mathbb{P}\big[\ \log_N(S_N(X)) \in [0, \log_N(p)]\ \big] = \log_N(p)$$

Wegen $N^{log_N(x)} = x$ und $N^0 = 1$ ergibt das Potenzieren mit der Basis N:

$$\mathbb{P}\big[\ S_N(X) \in [1, p]\ \big] = \log_N(p)$$

Somit folgt $S_N(X)$ dem Gesetz von Benford (Vgl. Definition 2.2). $\square$

3 Die Inverse Gammaverteilung

Im Folgenden wird kurz die Inverse Gammaverteilung vorgestellt und ihre Verbindung der Gammaverteilung aufgezeigt.

Definition 3.1 (Inverse Gammaverteilung)**.** Eine Zufallsvariable $X_{\alpha,\beta}$ mit Werten in $(0,\infty)$ folgt der Inversen Gammaverteilung mit den Parametern $\alpha,\beta > 0$, falls sie die Dichte

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(\frac{-\beta}{x}\right)$$

und die Verteilungsfunktion

$$F_{\alpha,\beta}(x) = \frac{1}{\Gamma(\alpha)} \int_{\frac{\beta}{x}}^{\infty} t^{\alpha-1} e^{-t} \, dt$$

hat.[F, Seite 112]

Bemerkung 3.2. Die Inverse Gammaverteilung ist wie die Inverse einer Gammaverteilung verteilt, denn sei X gammaverteilt mit den Parametern $a, b > 0$, dann ist $Y := \frac{1}{X}$ invers gammaverteilt mit den Parametern $\alpha = a$ und $\beta = \frac{1}{b}$. [C, Seite 2]

Mit dem Tranformationssatz für Dichten und der Dichte der Gammaverteilung X mit $a, b > 0$ mit der Form:

$$f_X(x) = \frac{b^{-a}}{\Gamma(a)} \exp\big(-\frac{x}{b}\big) x^{a-1}$$

gilt:

$$\begin{aligned} f_Y(y) &= f_X\big(\frac{1}{y}\big) \cdot \big|\frac{d}{dy} y^{-1}\big| = \\ &= \frac{b^{-a}}{\Gamma(a)} \exp\big(-\frac{1}{by}\big) y^{-a+1} \cdot y^{-2} = \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\big(-\frac{\beta}{y}\big) y^{-\alpha-1} \end{aligned}$$

Der Transformationssatz kann in [G, Seite 119] nachgeschlagen werden.

4 Die Annäherung der Inversen Gamma Verteilung an Benford's Law

Im folgenden Kapitel wird das Verhlten der Inverse Gammaverteilung und die Auswirkungen der Wahl der Parameter αund β auf die Abweichung zum Gesetz von Benford aufgezeigt.

Es soll zunächst Satz 2.5 auf die Inverse Gammaverteilung angewendet werden um einen Startpunkt für die Analyse zu erhalten,

Korollar 4.1. *Sei $X_{\alpha,\beta}$ invers Gamma-verteilt und sei F_N die Verteilungsfunktion von $\log_N(X_{\alpha,\beta}) \mod 1$. Dann folgt $X_{\alpha,\beta}$ dem Gesetz von Benford, falls für $[a,b] \subset [0,1]$ gilt:*

$$\mathbb{P}\Big[\log_N(X_{\alpha,\beta}) \mod 1 \in [a,b] \Big] = b - a$$

bzw. falls für alle $z \in [0,1]$ gilt:

$$\mathbb{P}\Big[\log_N(X_{\alpha,\beta}) \mod 1 \in [0,z] \Big] = F_N(z) = z$$

In den folgenden Abschnitten wird diese Definition verwendet um zu untersuchen für welche Parameter α und β die Abweichung von $F_N(z)$ zu z klein ist. Das wird erreicht, indem eine Reihenentwicklung von F'_N der Form 1+ Fehlerterm aufgestellt und integrier wird, um wieder zur Verteilungsfunktion zu gelangen. Hierbei sollte beachten werden, dass der Fehlerterm möglich genau berechnet werden kann.

Am Ende dieses Kapitels wird eine Schränke für den Restterm in der berechneten Reihenentwicklung aufgestellt und die Resultate wedren in mehreren Grafiken dargestellt.

4.1 Reihenentwicklung für F'_N

Bevor die Berechnung der Reihenentwicklung erläutert wird, braucht man noch eine Invarianzeigenschaft der Inversen Gammaverteilung um die Berechnung zu vereinfachen: die Abweichung vom Gesetz von Benford verändert sich nicht, wenn man den Parameter β mit Potenzen der Basis N multipliziert.

Lemma 4.2. *Für alle $\alpha, \beta > 0$ und $z \in [0,1]$ gilt nach [D, Seite 3]:*

$$\mathbb{P}\Big[\log_N S_N(X_{\alpha,\beta}) \le z \Big] = \mathbb{P}\Big[\log_N S_N(X_{\alpha,N\cdot\beta}) \le z \Big]$$

Beweis.

$$
\begin{aligned}
&\mathbb{P}\big[\log_N S_N(X_{\alpha,N\cdot\beta}) \le z\big] \\
&= \mathbb{P}\big[\log_N S_N(X_{\alpha,N\cdot\beta}) \in [0,z]\big] && (4.1)\\
&= \sum_{k=-\infty}^{\infty} \mathbb{P}\big[\log_N X_{\alpha,N\cdot\beta} \in [k, z+k]\big] && (4.2)\\
&= \sum_{k=-\infty}^{\infty} \mathbb{P}\big[X_{\alpha,N\cdot\beta} \in [N^k, N^{z+k}]\big] \\
&= \sum_{k=-\infty}^{\infty} \mathbb{P}\big[X_{\alpha,N\cdot\beta} \le N^{z+k}\big] - \mathbb{P}\big[X_{\alpha,N\cdot\beta} \le N^k\big] && (4.3)\\
&= \sum_{k=-\infty}^{\infty} \frac{1}{\Gamma(\alpha)} \int_{\frac{N\cdot\beta}{N^{z+k}}}^{\infty} t^{\alpha-1} e^{-t}\, dt - \frac{1}{\Gamma(\alpha)} \int_{\frac{N\cdot\beta}{N^{k}}}^{\infty} t^{\alpha-1} e^{-t}\, dt \\
&= \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \int_{\frac{N\cdot\beta}{N^{z+k}}}^{\frac{N\cdot\beta}{N^{k}}} t^{\alpha-1} e^{-t}\, dt && (4.4)\\
&= \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \int_{\frac{\beta}{N^{z+k-1}}}^{\frac{\beta}{N^{k-1}}} t^{\alpha-1} e^{-t}\, dt \\
&= \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \int_{\frac{\beta}{N^{z+k}}}^{\frac{\beta}{N^{k}}} t^{\alpha-1} e^{-t}\, dt && (4.5)\\
&= \mathbb{P}\big[\log_N S_N(X_{\alpha,\beta}) \le z\big]
\end{aligned}
$$

Da die Mantisse S_N nach Bemerkung 2.1 größer als 1 ist, sind die Aussagen: $S_N(X_{\alpha,N\cdot\beta}) \le z$ und $S_N(X_{\alpha,N\cdot\beta}) \in [0,z]$ äquivalent, wodurch man zu (4.1) kommt.

Für eine positive Zufallsvariable X kann wegen $i(X) \in \mathbb{Z}$ das Ereignis $\log_N S_N(X) \in [0,z]$ gleichgesetzt werden mit:

$$\log_N(S_N(X)) + i(X) \in \bigcup_{k=-\infty}^{\infty} [k, z+k]$$

$$\text{bzw. mit} \qquad \log_N(\underbrace{S_N(X)\cdot N^{i(X)}}_{=X}) \in \bigcup_{k=-\infty}^{\infty} [k, z+k]$$

Da die Intervalle $[k, z+k]$ alle disjunkt sind, kann die Wahrscheinlichkeit über alle Intervalle aufsummiert werden und man gelangt zu (4.2).

Es kann nun mit der Basis N potenziert und die Wahrscheinlichkeit, dass die Inverse Gammaverteilung in ein Intervall fällt, als Differenz dargestellt werden ((4.3)). In diese Darstellung kann die Verteilungsfunktion der Inversen Gammaverteilung eingesetzt und die Differenz durch die Intealität des Integrals gebilden werden ((4.4)).

In den Integralgrenzen kann das N jeweils im Nenner und im Zähler gekürzt werden. Durch einen Indexshift (da über alle ganzen Zahlen aufsummiert wird) erhält man (4.5).

Führt man nun laut [D, Seite 4] alle Schriit für die neuen Integrallgrenzen rückwärts aus, gelangt man zur Aussage. □

Ohne Beeinschränkung der Allgemeinheit kann ab hier angenommen werden, dass β keiner N ist.

Wie schon erwähnt, werden in diesem Abschnitt Bedingungen für $F_N(z) \sim z$ auf Bedingungen für $F_N'(z) \sim 1$ zrückgeführt.

Davor sollte man sich noch mit der Fourier-Transformierte für integrierbare Funktionen auf den Reellen Zahlen und der Poissonsche Summenformel vertraut machen. Diese sind im letzten Kaptiel, dem Appendix, unter 5.1 - 5.3 angeführt.

Satz 4.3. *Seien $\alpha, \beta \in \mathbb{R}, \alpha, \beta > 0$ und $N \in \mathbb{N}, N \geq 3$. Außerdem sei die Zufallsvariable $X_{\alpha,\beta}$ eine Inverse Gammverteilung und F_N die Verteilungsfunktion von $\log_N(X_{\alpha,\beta}) \mod 1$. Nach [D, Seite 4] ist dann F_N' gegeben durch:*

$$F_N'(z) = 1 + \frac{2}{\Gamma(\alpha)} \sum_{k=1}^{\infty} \mathrm{Re}\left(e^{2\pi i k(\log_N(\beta) - z)} \Gamma\left(\alpha - \frac{2\pi i k}{\ln(N)}\right)\right)$$

Beweis. Man betrachte zuerst die Verteilungsfunktion F_N an einer Stelle z. Wie im Beweis von Lemma 4.2 in (4.5) kann auch folgende Darstellung verwendet werden:

$$F_N(z) = \mathbb{P}[\log_N S_N(X_{\alpha,\beta}) \leq z] = \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \int_{\frac{\beta}{N^{z+k}}}^{\frac{\beta}{N^k}} t^{\alpha-1} e^{-t} \, dt \tag{4.6}$$

Dieser Ausdruck soll nun abgeleitet werden ableiten. Dazu zeigt man zuerst mit dem Weierstraß-Kriterium, dass die Reihe für $z \in [0,1]$ gleichmäßig konvergiert, um Differentation und Summe vertauschen zu dürfen. Das Weierstraß-Kriterium ist im Appendix unter 5.4 angeführt.

Man definiere zunächst eine Funktionen-Folge:

$$g_0(z) := \frac{1}{\Gamma(\alpha)} \int_{\frac{\beta}{N^z}}^{\beta} t^{\alpha-1} e^{-t} \, dt \qquad \text{und}$$

$$g_k(z) := \frac{1}{\Gamma(\alpha)} \left(\int_{\frac{\beta}{N^{z+k}}}^{\frac{\beta}{N^k}} t^{\alpha-1} e^{-t} \, dt + \int_{\frac{\beta}{N^{z-k}}}^{\frac{\beta}{N^{-k}}} t^{\alpha-1} e^{-t} \, dt \right) \qquad \text{für } k \geq 1$$

g_k ist somit die Summe aller Summanden aus (4.6) deren Indizes den Betrag k haben. Außerdem ist für ein festes $k \in \mathbb{Z}$ der Ausdruck $\frac{1}{N^{z+k}}$ streng monoton fallend in $z \in [0,1]$. Da auch der Integrand $t^{\alpha-1}\, e^{-t}$ positiv auf den Integrationsintervallen ist, ist g_k in $z \in [0,1]$ streng monton wachsend.

Es gilt also für alle $z \in [0,1]$ und alle $k \in \mathbb{N} \cup \{0\}$:

$$\|g_k(z)\|_\infty = g_k(1) := M_k$$

Zudem gilt:

$$\sum_{k=0}^{\infty} M_k = \sum_{k=0}^{\infty} g_k(1) \overset{(1)}{=} \mathbb{P}[\,\log_N S_N(X_{\alpha,\beta}) \le 1\,] \overset{(2)}{=} 1.$$

Gleichheit (1) kann analog wie im Beweis von Lemma 4.2 ausgeschrieben werden. Da die Mantisse S_N laut Definition kleiner N ist, ist der Ausdruck $\log_N S_N(X_{\alpha,\beta})$ immer kleiner 1, falls die Zufallsvariable $X_{\alpha,\beta}$ reelle Werte annimmt. Dadurch gilt Gleichheit (2).

Dadurch sind die Vorraussetzungen des Weierstraß-Kriteriums 5.4 erfüllt. Wegen der resultierenden gleichmäßigen Konvergenz auf $[0,1]$ von (4.6) darf die Differentiation in die Summe hinein gezogen werden. Die Ableitung des Integrals kann wie die eines Parameterintegrals behandelt werden. Die Vorgangsweise ist analog zu jener in [K, Seite 342]. Hierfür werden zunächst einige Funktionen definiert:

- das Parameterintegral:

$$I(z) := \int_{\frac{\beta}{N^{z+k}}}^{\frac{\beta}{N^k}} t^{\alpha-1}\, e^{-t}\, dt$$

- zwei stetig differenzierbare Funktionen, die die Integralgrenzen darstellen:

$$\alpha(z) := \frac{\beta}{N^{z+k}} \Big(= \beta \cdot e^{-(z+k)\ln N} \Big), \quad \beta(z) := \frac{\beta}{N^k}$$

- der stetige Integrand, für den die Ableitung nach der ersten Variable $\frac{\partial h}{\partial z}$ existiert:

$$h(z,t) := t^{\alpha-1}\, e^{-t}, \quad \text{mit} \quad \frac{\partial}{\partial z} h(z,t) = 0$$

- außerdem benötigt man eine Darstellung von I der Form $I(z) = (f \circ g)(z)$, dazu wähle man f und g folgendermaßen:

$$f(\xi,\eta,\zeta) := \int_\xi^\eta h(\zeta,t)\, dt \quad \text{und} \quad g(z) := (\alpha(z), \beta(z), z)^T$$

Die Voraussetzungen sind laut [K, Seite 342] ausreichend, dass die Ableitungen von f und g nach den einzelnen Variablen existieren und auch stetig sind (hier ist $x = (\xi, \eta, \zeta)$):

$$\begin{aligned} df(x) &= \Big(\frac{\partial f}{\partial \xi}(x), \frac{\partial f}{\partial \eta}(x), \frac{\partial f}{\partial \zeta}(x)\Big) = \Big(-h(\zeta,\xi), h(\zeta,\eta), \int_\xi^\eta \frac{\partial h}{\partial \zeta}(\zeta,t)\,dt\Big) \\ &= \Big(-\xi^{\alpha-1}e^{-\xi}, \zeta^{\alpha-1}e^{-\zeta}, 0\Big) \end{aligned}$$

$$dg(z) = (\alpha'(z), \beta'(z), 1) = \Big(-\beta\cdot e^{-(z+k)\ln N}\cdot \ln N, 0, 1\Big) = \Big(-\frac{\beta}{N^{z+k}}\cdot \ln N,\ 0,\ 1\Big)$$

Mit der Kettenregel (K), berechnet sich die Ableitung von I zu:

$$\begin{aligned} I'(z) &= d(f\circ g)(z) \overset{(K)}{=} df(g(z))dg(z) = \\ &= \Big(-\Big(\frac{\beta}{N^{z+k}}\Big)^{\alpha-1}\exp\Big(-\frac{\beta}{N^{z+k}}\Big), \Big(\frac{\beta}{N^k}\Big)^{\alpha-1}\exp\Big(\frac{\beta}{N^k}\Big), 0\Big)\cdot\begin{pmatrix} -\frac{\beta}{N^{z+k}}\ln N \\ 0 \\ 1\end{pmatrix} = \\ &= \Big(\frac{\beta}{N^{z+k}}\Big)^{\alpha}\exp\Big(-\frac{\beta}{N^{z+k}}\Big)\ln N \end{aligned}$$

Insgesamt erhält man:

$$F_N'(z) = \frac{1}{\Gamma(\alpha)}\sum_{k=-\infty}^{\infty}\Big(\frac{\beta}{N^{z+k}}\Big)^{\alpha}\exp\Big(-\frac{\beta}{N^{z+k}}\Big)\ln N$$

Als nächstes wird in [D, Seite 5] diese Darstellung mit Hilfe der Poissonschen Summenformel weiter umgeformt. Hierfür wird die Fourier-Transformierte von:

$$f_k(z) := \frac{1}{\Gamma(\alpha)}\Big(\frac{\beta}{N^{z+k}}\Big)^{\alpha}\exp\Big(-\frac{\beta}{N^{z+k}}\Big)\ln N$$

nach 5.1 berechnet:

$$\hat{f}_\xi(z) = \int_{-\infty}^{\infty}\frac{1}{\Gamma(\alpha)}\Big(\frac{\beta}{N^{z+x}}\Big)^{\alpha}\exp\Big(-\frac{\beta}{N^{z+x}}\Big)\ln N\ \exp(-2\pi i x\xi)\,dx$$

Einsetzten in die Poissonsche Summenformel 5.3 bringt:

$$\begin{aligned} F_N'(z) &= \sum_{k=-\infty}^{\infty} f_k(z) = \sum_{k=-\infty}^{\infty}\hat{f}_k(z) = \\ &= \frac{1}{\Gamma(\alpha)}\sum_{k=-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(\frac{\beta}{N^{z+x}}\Big)^{\alpha}\exp\Big(-\frac{\beta}{N^{z+x}}\Big)\ln N\ \exp(-2\pi i x k)\,dx \end{aligned}$$

Nun wird mit

$$u = \frac{\beta}{N^{z+x}} = \beta e^{-(z+x)\ln N} \qquad \text{und}$$

$$\frac{du}{dx} = -\beta e^{-(z+x)\ln N}\ln N = -\frac{\beta\ln N}{N^{z+x}} = -u\ln N \qquad \text{bzw.} \qquad dx = -\frac{1}{u\ln N}\,du\,.$$

substituiert und x wird geschrieben als:

$$\begin{aligned} u &= \frac{\beta}{N^{z+x}} \\ \Leftrightarrow \; N^{z+x} &= \frac{\beta}{u} \\ \Leftrightarrow \; z + x &= \log_N \frac{\beta}{u} \\ \Leftrightarrow \qquad x &= \log_N \frac{\beta}{u} - \log_N N^z \\ \Leftrightarrow \qquad x &= \log_N \frac{\beta}{uN^z} = \frac{\ln \frac{\beta}{uN^z}}{\ln N} \end{aligned}$$

Im letzten Schritt wird noch ein Basiswechsel für den Logarithmus vorgenommen:

$$\log_b a = \frac{\log_c a}{\log_c b} \quad \text{für}\, a, b, c > 0$$

Durch die Subustitution werden hier auch die Integralgrenzen geändert und nach weiteren Umformungen erhält man:

$$\begin{aligned} F_N'(z) &= \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \int_{\infty}^{0} u^{\alpha} \exp(-u) \ln N \, \exp(-2\pi i x k) \Big(-\frac{1}{u \ln N} \Big) \, du = \\ &= \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \int_{0}^{\infty} u^{\alpha-1} \exp(-u) \exp\left(-\frac{2\pi i k}{\ln N} \cdot \ln\Big(\frac{\beta}{uN^z}\Big)\right) du = \\ &= \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \int_{0}^{\infty} u^{\alpha-1} \exp(-u) \Big(\frac{\beta}{uN^z}\Big)^{-\frac{2\pi i k}{\ln N}} du = \\ &= \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \Big(\frac{\beta}{N^z}\Big)^{-\frac{2\pi i k}{\ln N}} \int_{0}^{\infty} u^{\alpha-1+\frac{2\pi i k}{\ln N}} \, e^{-u} \, du = \\ &= \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \Big(\frac{\beta}{N^z}\Big)^{-\frac{2\pi i k}{\ln N}} \Gamma\Big(\alpha + \frac{2\pi i k}{\ln N}\Big) \\ &= \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{2\pi i k}{\ln N} \cdot \ln\Big(\frac{\beta}{N^z}\Big)\right) \Gamma\Big(\alpha + \frac{2\pi i k}{\ln N}\Big) \end{aligned} \tag{4.7}$$

Hier wird in der vorletzten Zeile mit $\Gamma(x) = \int_0^\infty u^{x-1} \, e^{-u} \, du$ die Gammafunktion bezeichnet.

Die Summe wird jetzt in jenen Teil mit positven Index, negativen Index und in den Summanden für $k = 0$ aufgeteilt. Da der Summand für $k = 0$ gleich 1 ist, und nach einem erneuten Basiswechsel innerhalb der Exponentialfunktion kommt man zu:

$$\begin{aligned} F_N'(z) = 1 + \frac{1}{\Gamma(\alpha)} \Bigg[&\sum_{k=0}^{\infty} \exp\left(-2\pi i k \cdot \log_N \Big(\frac{\beta}{N^z}\Big)\right) \Gamma\Big(\alpha + \frac{2\pi i k}{\ln N}\Big) + \\ &+ \sum_{k=0}^{\infty} \exp\left(2\pi i k \cdot \log_N \Big(\frac{\beta}{N^z}\Big)\right) \Gamma\Big(\alpha - \frac{2\pi i k}{\ln N}\Big) \Bigg] \end{aligned}$$

Im nachfolgenden und letzten Schritt verwende man die Eigenschaft der Gammafunktion $\overline{\Gamma(z)} = \Gamma(\overline{z})$ und der Exponentialfunktion $\overline{\exp(z)} = \exp(\overline{z})$ für komplexe Zahlen $z := a + ib$ mit reellen a, b.

Außerdem, da für komplex konjugierte Zahlen mit reellen a_i, b_i die Rechenregeln:

$$\overline{(a_1 + ib_1)} \cdot \overline{(a_2 + ib_2)} = \overline{(a_1 + ib_1)(a_2 + ib_2)} \tag{4.8}$$

$$(a + ib) + \overline{(a + ib)} = 2\,\mathrm{Re}(a + ib) \tag{4.9}$$

gelten, hat man schlussendlich wie in [D, Seite 6]:

$$\begin{aligned}
F_N'(z) = 1 + \frac{1}{\Gamma(\alpha)}\Bigg[&\sum_{k=0}^{\infty} \overline{\exp\Big(2\pi ik \cdot \log_N\Big(\frac{\beta}{N^z}\Big)\Big)}\,\overline{\Gamma\Big(\alpha - \frac{2\pi ik}{\ln N}\Big)} + \\
&+ \sum_{k=0}^{\infty} \exp\Big(2\pi ik \cdot \log_N\Big(\frac{\beta}{N^z}\Big)\Big)\Gamma\Big(\alpha - \frac{2\pi ik}{\ln N}\Big)\Bigg] = \\
\overset{(2.12)}{\underset{(2.13)}{=}} 1 + \frac{2}{\Gamma(\alpha)} \sum_{k=0}^{\infty} \mathrm{Re}&\Bigg(\exp\Big(2\pi ik \cdot \log_N\Big(\frac{\beta}{N^z}\Big)\Big)\Gamma\Big(\alpha - \frac{2\pi ik}{\ln N}\Big)\Bigg) = \\
= 1 + \frac{2}{\Gamma(\alpha)} \sum_{k=0}^{\infty} \mathrm{Re}&\Bigg(\exp\Big(2\pi ik \cdot (\log_N(\beta) - z)\Big)\Gamma\Big(\alpha - \frac{2\pi ik}{\ln N}\Big)\Bigg) = \\
= 1 + \frac{2}{\Gamma(\alpha)} \sum_{k=0}^{\infty} \mathrm{Re}&\Bigg(e^{2\pi ik\cdot(\log_N(\beta) - z)}\,\Gamma\Big(\alpha - \frac{2\pi ik}{\ln N}\Big)\Bigg)
\end{aligned} \tag{4.10}$$

□

4.2 Fehlerabschätzung

Im vorherigen Abschnitt wurde eine Darstellung der Ableitung der Verteilungsfunktion F_N' der Zufallsvariable $\log_N X_{\alpha,\beta} \mod 1$, wobei $X_{\alpha,\beta}$ invers-gammaverteilt ist, hergeleitet. Nun soll noch eine Abschätzung für die Summe in dieser Darstellung aufgestellt werden.

Satz 4.4 (Fehlerabschätzung). *Sei F_N' die Verteilungsfunktion von $\log_N X_{\alpha,\beta} \mod 1$ mit einer invers gamma-verteilten Zufallsvariable $X_{\alpha,\beta}$ mit beliebigen $\alpha, \beta > 0$. F_N' habe die Darstellung wie in (4.7):*

$$F_N'(z) = \frac{1}{\Gamma(\alpha)} \sum_{k=-\infty}^{\infty} \exp\Big(-\frac{2\pi ik}{\ln N} \cdot \ln\Big(\frac{\beta}{N^z}\Big)\Big)\Gamma\Big(\alpha + \frac{2\pi ik}{\ln N}\Big)$$

Weiters sei mit E_M ('Error') jener Teil der Summe notiert, dessen Indizes einen Absolutbetrag größer gleich M hat:

$$E_M(z) = \frac{1}{\Gamma(\alpha)} \sum_{|k| \geq M} \exp\Big(-\frac{2\pi ik}{\ln N} \cdot \ln\Big(\frac{\beta}{N^z}\Big)\Big)\Gamma\Big(\alpha + \frac{2\pi ik}{\ln N}\Big)$$

Dann gilt:

(1) $E_M(z)$ für $z \in [0,1]$ kann abgeschätzt werden durch:

$$|E_M(z)| \leq \frac{N^{\alpha(1-z)}\beta}{\Gamma(\alpha)}\left(\int_{N^M}^{\infty} e^{-x}\, x^{\alpha-1}\, dx + \frac{1}{\alpha} N^{-M\alpha}\right)$$

(2) Für alle $z \in [0,1]$ kann E_M durch folgende Konstante abgeschätzt werden:

$$|E_M| \leq \frac{N^{\alpha}\,\beta}{\Gamma(\alpha)}\left(\int_{N^M}^{\infty} e{-}x\, x^{\alpha-1}\, dx + \frac{1}{\alpha}\, N^{-M\alpha}\right)$$

(3) Für alle $\epsilon > 0$ und $M > \max(\alpha+1, -\log_N(\frac{\epsilon\cdot\Gamma(\alpha)}{2N^{\alpha}\beta}))$ gilt: $|E_M(z)| \leq \epsilon$ *[D, Seite 6]*

Beweis von (1). Zunächst definiere man sich mit $\gamma(z) := \ln\left(\frac{\beta}{N^z}\right)$ eine Darstellung für E_1:

$$E_1(z) = \frac{1}{\Gamma(\alpha)} \sum_{|k|\geq 1} \exp\left(-\frac{2\pi i k}{\ln N}\cdot\gamma(z)\right)\Gamma\left(\alpha + \frac{2\pi i k}{\ln N}\right)$$

Der Faktor

$$\Gamma\left(\alpha + \frac{2\pi i k}{\ln N}\right) = \int_0^{\infty} x^{-1+\alpha+\frac{2\pi i k}{\ln N}}\, e^{-x}\, dx$$

kann mit $x = e^{-u} (\Leftrightarrow -\ln x = u)$ und $\frac{du}{dx} = -\frac{1}{x} = -e^{u}$ substituiert werden, wodurch sich folgendes ergibt:

$$\begin{aligned}\Gamma\left(\alpha + \frac{2\pi i k}{\ln N}\right) &= \int_{\infty}^{-\infty} e^{-u^{(-1+\alpha+\frac{2\pi i k}{\ln N})}}\, e^{-e^{-u}}\, (-e^{-u})\, du = \\ &= \int_{-\infty}^{\infty} e^{-u\cdot(-1+\alpha+\frac{2\pi i k}{\ln N})}\, e^{-e^{-u}}\, e^{-u}\, du = \\ &= \int_{-\infty}^{\infty} e^{-u(\alpha+\frac{2\pi i k}{\ln N})}\, e^{-e^{-u}}\, du \\ &= \int_{-\infty}^{\infty} e^{-e^{-u}}\, e^{-u\alpha}\, e^{-2\pi i u\frac{k}{\ln N}}\, du \end{aligned} \tag{4.11}$$

Vergleicht man nun (4.11) mit der Definition der Fouriertransformierten aus 5.1, kann man (4.11) auch schreiben als die Fouriertransformierte von $s(x) = e^{-e^{-x}}\, e^{-x\alpha}$ an der Stelle $\xi = \frac{k}{\ln N}$. Diese Fouriertransformierte sei als $\hat{s}\left(\frac{k}{\ln N}\right)$ notiert.

Als Zwischenschritt erhält man folgende Darstellung für E_1:

$$E_1(z) = \frac{1}{\Gamma(\alpha)} \sum_{|k|\geq 1} \exp\left(-\frac{2\pi i k}{\ln N}\cdot\gamma(z)\right)\cdot\hat{s}\left(\frac{k}{\ln N}\right) \tag{4.12}$$

Als nächstes wird mit einer ßcaling and frequency propertity of the fouriertransformationünd der Possionsches Summenformel die Darstellung (4.12) weiter umgeformt:

Sei $s_0 \in L^1(\mathbb{R}), D > 0$ und $v \in \mathbb{R}$ und sei $g(x) = s_0(Dx+v)$, dann gilt mit Substitution mit $u = Dx + v \Leftrightarrow \frac{u-v}{D} = x$ und $\frac{du}{dx} = D$ folgender Zusammenhang für die Fouriertransformierten von g und s:

$$\begin{aligned}\hat{g}(\xi) &= \int_{-\infty}^{\infty} g(x)\, e^{-2\pi i x\xi}\, dx = \int_{-\infty}^{\infty} s_0(Dx+v)\, e^{-2\pi i x\xi}\, dx = \\ &= \int_{-\infty}^{\infty} s_0(u) \exp\Big(-2\pi i \,\frac{u-v}{D}\,\xi\Big)\, \frac{1}{D}\, du = \\ &= \frac{1}{D} \exp\Big(\frac{2\pi i v\xi}{D}\Big) \int_{-\infty}^{\infty} s_0(u) \exp\Big(-\frac{2\pi i u\xi}{D}\Big)\, du = \\ &= \frac{1}{D} \exp\Big(\frac{2\pi i v\xi}{D}\Big)\, \hat{s_0}\Big(\frac{\xi}{D}\Big)\end{aligned} \tag{4.13}$$

Mit der Poissonschen Summenformel gilt nun für $s_0(Dx+v)$ und (4.13):

$$D \cdot \sum_{n\in\mathbb{Z}} s_0(Dn+v) = \sum_{n\in\mathbb{Z}} \exp\Big(\frac{2\pi i v\xi}{D}\Big)\, \hat{s_0}\Big(\frac{\xi}{D}\Big) \tag{4.14}$$

Für $s_0(x) = s(x) = e^{-e^{-x}}, D = \ln N, v = -\gamma(z)$ und $\xi \,\hat{=}\, k$ (ξ entspricht k) kann (4.12) weiter abgeschätzt werden:

$$\begin{aligned}E_1(z) &\overset{(4.12)}{=} \frac{1}{\Gamma(\alpha)} \sum_{|k|\geq 1} \exp\Big(-\frac{2\pi i k}{\ln N}\cdot\gamma(z)\Big)\cdot \hat{s}\Big(\frac{k}{\ln N}\Big) = \\ &= \left(\frac{1}{\Gamma(\alpha)} \sum_{k\in\mathbb{Z}} \exp\Big(-\frac{2\pi i k v}{D}\Big)\cdot \hat{s}\Big(\frac{k}{D}\Big)\right) - \frac{\hat{s}(0)}{\Gamma(\alpha)} = \\ &\leq \frac{1}{\Gamma(\alpha)} \sum_{k\in\mathbb{Z}} \exp\Big(-\frac{2\pi i k v}{D}\Big)\cdot \hat{s}\Big(\frac{k}{D}\Big) = \\ &\overset{(4.14)}{=} \frac{D}{\Gamma(\alpha)} \sum_{k\in\mathbb{Z}} s\,(Dk+v) =\end{aligned} \tag{4.15}$$

Addiert man den Summanden für $k = 0$ und zieht ihn gleich wieder ab, entsteht Gleichung (4.15). Zusätzlich, da s nicht negativ ist, ist auch $\hat{s}$ nicht negativ. Somit kann der Term $-\frac{\hat{s}(0)}{\Gamma(\alpha)}$ weggelassen werden und man erhält eine Abschätzung nach oben.

In diese Abschätzung wird s eingesetzen und alle Terme, die unabhängig von k sind, werden vor die Summe gezogen:

$$\begin{aligned}&= \frac{D}{\Gamma(\alpha)} \sum_{k\in\mathbb{Z}} \exp(-e^{\gamma(z)-k\ln N})\, e^{\alpha(\gamma(z)-k\ln N)} = \\ &= \frac{e^{\alpha\gamma(z)}\ln N}{\Gamma(\alpha)} \sum_{k\in\mathbb{Z}} \exp(-e^{\gamma(z)}\, e^{-\ln N^k})\, e^{-\alpha k\ln N} = \\ &= \frac{e^{\alpha\gamma(z)}\ln N}{\Gamma(\alpha)} \sum_{k\in\mathbb{Z}} \exp(-\beta N^{-z} N^{-k})\, e^{-\alpha k \ln N} =\end{aligned}$$

$$\leq \frac{e^{\alpha\gamma(z)}\ln N}{\Gamma(\alpha)} \sum_{k\in\mathbb{Z}} \exp(\,-\beta N^{-k-1}\,)\, e^{-\alpha k \ln N} = \tag{4.16}$$
$$= \frac{e^{\alpha\gamma(z)}\ln N}{\Gamma(\alpha)} \sum_{k\in\mathbb{Z}} \exp(\,-\beta\, e^{-(k+1)\ln N}\,)\, e^{-\alpha k \ln N}$$

Durch die Annahme $z \leq 1$ gilt $N^z \leq N$ und $-N^{-z} \leq -N^{-1}$, womit man eine weitere Abschätzung in (4.16) erhält.

Analog berechnet sich eine Abschätzung für E_M zu:

$$E_M(z) \;\leq\; \frac{e^{\alpha\gamma(z)}\ln N}{\Gamma(\alpha)} \sum_{|k|\geq M} \exp(-\beta\, e^{-(k+1)\ln N})\, e^{-\alpha k \ln N} \tag{4.17}$$

Die Summe kann jetzt als Untersumme interpretiert und mit einem Integral nach oben abgeschätzt werden.

Mit der Substitution $x = e^{-(k+1)\ln N}$ bzw. $dx = -\ln N \cdot e^{-(k+1)\ln N} dk = -\ln N\, x\, dk$ (wodurch sind auch die Integralgrenzen c zu $e^{-(c+1)\ln N}$ ändern) erhält man:

$$\begin{aligned}
E_M \;&\leq\; \frac{e^{\alpha\gamma(z)}}{\Gamma(\alpha)}\Bigg(\int_{N^M}^{\infty} \exp(-\beta\, e^{-(k+1)\ln N})\, x^{-1}\, e^{-\alpha k \ln N}\, dx\; + \\
&\qquad + \int_0^{N^{-M}} \exp(-\beta\, e^{-(k+1)\ln N})\, x^{-1}\, e^{-\alpha k \ln N}\, dx\Bigg) && (4.18)\\
&= \frac{e^{\alpha\gamma(z)}}{\Gamma(\alpha)}\Bigg(\int_{N^M}^{\infty} \exp(-\beta x)\, x^{-1}\, e^{-\alpha(k+1)\ln N}\, e^{\alpha \ln N}\, dx\; + \\
&\qquad + \int_0^{N^{-M}} \exp(-\beta\, x)\, x^{-1}\, e^{-\alpha(k+1)\ln N}\, e^{\alpha \ln N} dx\Bigg) = && (4.19)\\
&= \frac{N^\alpha\, e^{\alpha\gamma(z)}}{\Gamma(\alpha)}\Bigg(\int_{N^M}^{\infty} e^{-\beta x}\, x^{\alpha-1}\, dx \;+ \int_0^{N^{-M}} e^{-\beta\, x}\, x^{\alpha-1}\, dx\Bigg) = \\
&\leq \frac{N^\alpha\, e^{\alpha\gamma(z)}}{\Gamma(\alpha)}\Bigg(\int_{N^M}^{\infty} e^{-x}\, x^{\alpha-1}\, dx \;+ \int_0^{N^{-M}} e^{-x}\, x^{\alpha-1}\, dx\Bigg) = && (4.20)
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{N^\alpha\, e^{\alpha \ln(\frac{\beta}{N^z})}}{\Gamma(\alpha)}\Bigg(\int_{N^M}^{\infty} e^{-x}\, x^{\alpha-1}\, dx \;+ \int_0^{N^{-M}} x^{\alpha-1}\, dx\Bigg) = && (4.21)\\
&= \frac{N^{\alpha(1-z)}\,\beta^\alpha}{\Gamma(\alpha)}\Bigg(\int_{N^M}^{\infty} e^{-x}\, x^{\alpha-1}\, dx + \Big[\frac{1}{\alpha}\, x^\alpha\Big]_{x=0}^{N^{-M}}\Bigg) = \\
&= \frac{N^{\alpha(1-z)}\,\beta^\alpha}{\Gamma(\alpha)}\Bigg(\int_{N^M}^{\infty} e^{-x}\, x^{\alpha-1}\, dx + \frac{1}{\alpha}\, N^{-M\alpha}\Bigg)
\end{aligned}$$

Durch den Variablenwechsel kürzt sich das $\ln N$ mit jenem aus dem Bruch aus (4.17) und das Minus gleicht jenes Minus aus, das entsteht, wenn die Intervallgrenzen vertauscht werden.

Um das k aus dem letzten Term in (4.18) auch noch zu substituieren, erweite man mit dem Bruch $e^{\alpha \ln N}/e^{\alpha \ln N}$, wodurch sich $x^\alpha \cdot N^\alpha$ (Vgl. (4.19)) ergibt.

Mit der Annahme $\beta \geq 1$, kann das β in der Exponentialfunktion bzw. die Exponentialfunktion im zweiten Integral, für eine Abschätzung nach oben, weggelassen werden((4.20) bzw. (4.21)).

Falls $\alpha \leq 1$ ist, ergibt sich noch eine weiter Abschätzung:

$$E_M(z) \;\leq\; \frac{N^{\alpha(1-z)}\,\beta}{\Gamma(\alpha)} \left(\int_{N^M}^{\infty} e^{-x}\, x^{\alpha-1}\, dx + \frac{1}{\alpha}\, N^{-M\alpha} \right)$$

□

Beweis von (2). Um die Konstante für alle $z \in [0,1]$ zu erhalten, maximiere man die Abschätzung aus (1) in z. Dafür die die Monotonie von $N^{\alpha(1-z)}$ wichtig.

Die Ableitung:

$$\frac{d}{dz}\, N^{\alpha(1-z)}\, \beta = \frac{d}{dz}\, e^{\alpha(1-z)\ln N}\, \beta = -\ln N \cdot e^{\alpha(1-z)\ln N}\, \beta = -\ln N \cdot N^{-\alpha(1-z)}\, \beta$$

ist für $z \in [0,1]$ negativ, wodurch $N^{\alpha(1-z)}$ monoton fallend ist. Deswegen liegt das Maximum bei $z = 0$.

E_M kann man daher abschätzen durch:

$$|\,E_M\,| \;\leq\; \frac{N^\alpha\, \beta^\alpha}{\Gamma(\alpha)} \left(\int_{N^M}^{\infty} e^{-x}\, x^{\alpha-1}\, dx + \frac{1}{\alpha}\, N^{-M\alpha} \right)$$

und für $\alpha \leq 1$ gilt sogar:

$$|\,E_M\,| \;\leq\; \frac{N^\alpha\, \beta}{\Gamma(\alpha)} \left(\int_{N^M}^{\infty} e^{-x}\, x^{\alpha-1}\, dx + \frac{1}{\alpha}\, N^{-M\alpha} \right)$$

□

Beweis von (3). Um Aussage (3) zu zeigen, wird eine weitere Abschätzung für die Schranke aus (2) berechnet. Insbeseondere soll das Integral über $e^{-x}\, x^{\alpha-1}$ abgeschätzt werden.

Fall 1: Sei $M = \alpha + 1$, dann folgt aus $N \geq 3$ die Ungleichung:

$$N^{\alpha+1} = N^M > e^{\alpha+1}$$

Fall 2: Sei $M = -\log_N(\frac{\epsilon \cdot \Gamma(\alpha)}{2N^\alpha \beta})$, dann gilt wegen $M \geq \alpha + 1$:

$$N^M \geq N^{\alpha+1} > e^{\alpha+1}$$

Also ist $N^M > e^{\alpha+1}$ für $N \geq 3$.

Wenn man nun auf beiden Seiten den Logarithmus anwendet, ergibt das unmittelbar:

$$\ln N^M > \alpha + 1 \qquad \text{bzw.} \qquad \frac{1}{\ln N^M} < \frac{1}{\alpha+1}$$

Da alle Ausdrücke positiv sind, können sie zusammengesetzen werden zu:

$$\frac{N^M}{\ln N^M} > \frac{e^{\alpha+1}}{\alpha+1} > \alpha + 1$$

Betrachtet man nun von $\frac{x}{\ln x}$ die Ableitung: $\frac{\ln x - 1}{\ln^2 x}$ ist ersichtlich, dass diese positiv für $x > \exp(1)$ ist, wodurch $\frac{x}{\ln x}$ monoton steigend für $x > \exp(1)$ ist. Somit haben wir:

$$\frac{x}{\ln x} > \alpha + 1 \quad \text{für} \quad x \geq N^M.$$

Durch Umformungen erhalten wir:

$$\begin{aligned} & x > (\alpha+1)\ln x \\ \Leftrightarrow\quad & x > \ln x^{\alpha+1} \\ \Leftrightarrow\quad & -x < \ln x^{-(\alpha+1)} \\ \Leftrightarrow\quad & e^{-x} < x^{-(\alpha+1)} \\ \Leftrightarrow\quad & e^{-x}\, x^{\alpha-1} < x^{-2} \end{aligned}$$

Außerdem gilt nach [D, Seite 8f.] wegen der Wahl von M:

$$\frac{1}{\alpha} N^{-M\alpha} + N^{-M} < N^{-M} + N^{-M} = 2N^{-M} < 2N^{\log_N \frac{\epsilon \cdot \Gamma(\alpha)}{2N^\alpha \beta}} = \frac{\epsilon \cdot \Gamma(\alpha)}{N^\alpha \beta}$$

Insgesamt kommt man zu der Abschätzung:

$$\begin{aligned} |E_M(z)| \quad &< \quad \frac{N^\alpha \beta}{\Gamma(\alpha)} \left(\frac{1}{\alpha} N{-}M\alpha + \int_{N^M}^{\infty} \frac{1}{x^2}\, dx \right) \\ &< \quad \frac{N^\alpha \beta}{\Gamma(\alpha)} \left(\frac{1}{\alpha} N^{-M\alpha} + N^{-M} \right) \\ &< \quad \frac{N^\alpha \beta}{\Gamma(\alpha)} \cdot \frac{\epsilon \cdot \Gamma(\alpha)}{N^\alpha \beta} \quad = \quad \epsilon \end{aligned}$$

□

4.3 Intepretation und Graphiken

Mit diesen Resultaten möchte man nun die Inverse Gamma-Verteilung mit dem Gesetz von Benford vergleichen. Hierfür wird die Abteilung der Verteilungsfunktion aus Satz 4.3 von 0 bis z integriert und erhält wie in [D, Seite 9]:

$$F_N(z) = z + \frac{1}{\Gamma(\alpha)} \sum_{|k| \geq 1} e^{-2\pi i k \log_N \beta} \frac{1}{2\pi i k} \left(e^{2\pi i k z} - 1 \right) \Gamma\!\left(\alpha + \frac{2\pi i k}{\ln N}\right)$$

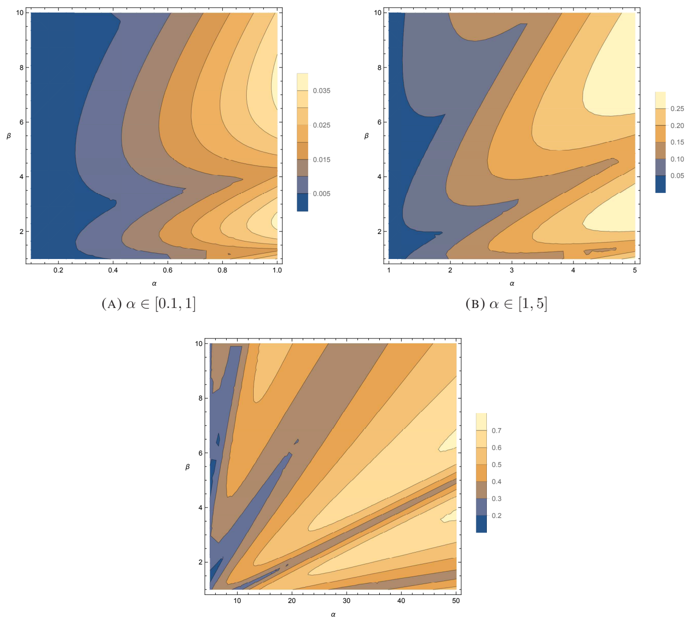
Wählt man nun ein $\epsilon > 0$, kann mit Hilfe von Satz 4.4 (3) der Wert von $|F'_N(z) - 1|$ mit einer Fehlerschranke der Größe ϵ bestimmt werden. Nach dem Mittelwertsatz der Differentialrechnung kann man ebenfalls $|F_N(z) - z|$ mit einer Genauigkeit der Größe ϵ bestimmen [D, Seite 9].

Insbesondere ist aus Satz 4.4 ersichtlich, dass für eine Berechnung mit einer Genuigkeit der Größe ϵ nur die ersten M Summanden berücksichtigt werden müssen.

In der nachstehenden Abbildung 4.1 wird $\max\limits_{z \in [0,1]} |F_{10}(z) - z|$ in Abhängigkeit von α und β mit einer Fehlerschranke von $\epsilon = 0.001$ visualisiert. In der Legende ist die Bedeutung der Schattierung zu finden: je dünkler die Fläche, desto kleiner ist die Abweichung von $F_{10}(z)$ zu z für die entsprechenden α und β.

Während sich die drei Teilgraphiken die y-Achse mit $\beta \in [1, 10]$ teilen unterscheiden sie sich in der x-Achse. Diese repräsentiert die verwendeten α auf den Intervallen $[0.1, 1]$, $[1, 5]$ und $[5, 50]$.

Aufgrund der Schattierung wird eine Tendenz zwischen der Abweichung von $F_{10}(z)$ zu z und den beiden verwendeten Paramater ersichtlich. Hier gibt es zwei wesentliche Merkmale, die auffällig sind. Einerseits werden die Flächen heller, je größer das α wird und andererseits bleibt die Schattierung ungefähr gleich, wenn β verändert wird. Für die Basis 10 lässt das auf eine übermäßige Abhängigkeit des Parameters α und einen geringeren Einfluss von β schließen.



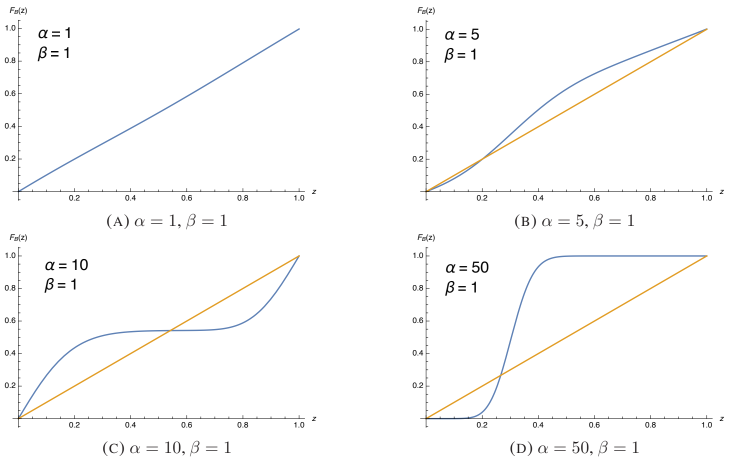
(A) $\alpha \in [0.1, 1]$

(B) $\alpha \in [1, 5]$

(C) $\alpha \in [5, 50]$

Abbildung 4.1: Abweichung $\max_{z \in [0,1]} |F_{10}(z) - z|$ in Abhängigkeit von α, β, [D, Seite 10]

In den nächsten beiden Abbildungen sind nun $F_{10}(z)$ in blau und $f(z) = z$ in orange nebeneinander geplotet. In Abbildung 4.2 wird $\beta = 1$ festgehalten und α variiert in $\{1, 5, 10, 50\}$. Auch hier zeigt sich die Abhängigkeit des Parameters α: die Verteilungsfunktion F_{10} weist größere Abweichungen von der gewünschten Benford Eigenschaft $f(z) = z$ auf, je größer α wird. [D, Seite 12]

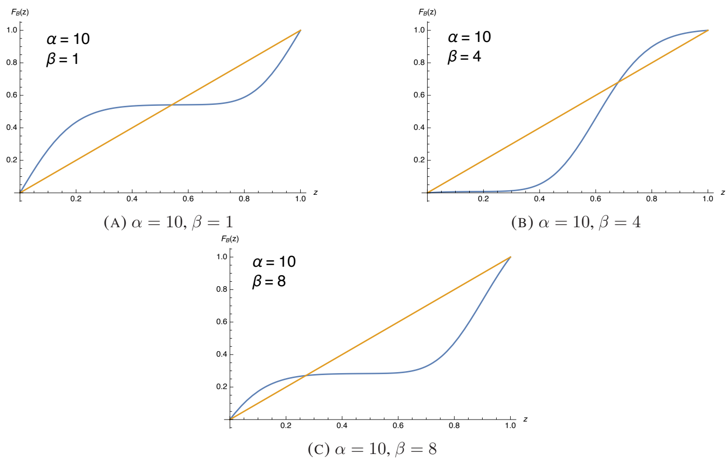


(A) $\alpha = 1, \beta = 1$ (B) $\alpha = 5, \beta = 1$

(C) $\alpha = 10, \beta = 1$ (D) $\alpha = 50, \beta = 1$

Abbildung 4.2: Verläufe von $F_{10}(z)$ und $f(z) = z$ (in orange) mit festen β, [D, Seite 11]

Die umgekehrte Situation ist in Abbildung 4.3 dargestellt. Hier wird $\alpha = 10$ festgehalten und β variiert in $\{1, 4, 8\}$. Genau wie in Abbildung 4.1 bleibt die Abweichung von $F_{10}(z)$ zu $f(z) = z$ in derselben Größenordnung, wenn β erhöht wird [D, Seite 12].



(A) $\alpha = 10, \beta = 1$ (B) $\alpha = 10, \beta = 4$

(C) $\alpha = 10, \beta = 8$

Abbildung 4.3: Verläufe von $F_{10}(z)$ und $f(z) = z$ (in orange) mit festen α ,[D, Seite 12]

5 Appendix

Definition 5.1 (Fourier-Transformierte)**.** Sei $f \in L^1(\mathbb{R})$. Die Fourier-Transformierte $\hat{f}$ von f ist dann gegeben durch:

$$\hat{f}(\xi) := \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i x \xi}\, dx$$

Die Funktion f hat dann die Darstellung:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi)\, e^{2\pi i x \xi}\, d\xi$$

[S, Seite 129ff.]

Die Fourier-Transformierte kann auch mit einem Vorfaktor $\frac{1}{\sqrt{2\pi}}$ definiert werden, dann wird die Exponentialfunktion zu: e^{-inx}.

Definition 5.2 (Schwartzraum über $\mathbb{R}$)**.** Der Schwartzraum $\mathcal{S}$ über $\mathbb{R}$ ist ein linearer Unteraum von $L^p(\mathbb{R})$ für $1 \leq p \leq \infty$ und besteht aus allen beliebig oft differenzierbaren Funktionen f, sodass alle Ableitungen $f^{(k)}$ fallend im folgenden Sinne sind:

$$\sup_{x \in \mathbb{R}} |x|^l |f^{(k)}(x)| < \infty \qquad \forall l, k \geq 0$$

[S, Seite 135]

Satz 5.3 (Possionsche Summenformel)**.** *Sei $f \in \mathcal{S}$, dann gilt:*

$$\sum_{n=-\infty}^{\infty} f(x+n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)\, e^{2\pi i n x}$$

Insbesondere gilt für $x = 0$:

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)$$

[S, Seite 154]

Satz 5.4 (Weierstraß Kriterium)**.** *Sei $(g_k)_{k \in \mathbb{N}}$ eine Folge beschränkter Funktionen, die in die komplexen Zahlen $\mathbb{C}$ abbilden. Wenn es eine Folge von Konstanten $(M_k)_{k \in \mathbb{N}} \in [0, \infty)$ gibt, sodass die Reihe $\sum_{k=1}^{\infty}$ konvergiert und $\|g_k\|_\infty \leq M_k$ für alle $k \in \mathbb{N}$, dann konvergiert $\sum_{k=1}^{\infty} g_k$ gleichmäßig. [K, Seite 193]*

Abbildungsverzeichnis

2.1 Verteilung der führenden Ziffer zur Basis 10 3

4.1 Abweichung $\max_{z\in[0,1]} |F_{10}(z) - z|$ in Anhängigkeit von α, β 19
4.2 Verlauf von $F_{10}(z)$ und z mit festem β 20
4.3 Verlauf von $F_{10}(z)$ und z mit festem α 20

Literaturverzeichnis

[C] John D. Cook: *Inverse Gamma Distribution*,3.10.2008, https://www.johndcook.com/inverse_ gamma.pdf, zuletzt besucht: 28.2.2021

[D] Rebecca F. Durst, Chi Huynh, Adam Lott,Steven J. Miller,Eyvindur A. Palsson,Wouter Touw,Gert Vriend: *The Inverse Gamma Distribution and Benford's Law*, 6.1.2018, https://arxiv.org/abs/1609.04106, zuletzt besucht: 28.2.2021

[F] Catherine Forbes, Merran Evans, Nicholas Hastings, Brian Peacock: *Statistical Distributions*, John Wiley & Sons, Inc., Hoboken, New Jersey, (2011),vierte Auflage, online Zugang: http://personalpages.to.infn.it/ zaninett/pdf/statistical-distributions.pdf, zuletzt besucht: 28.2.2021

[G] Karl Grill: *Maß- und Wahrscheinlichkeitstheorie*, 23.1.2018, Institut für Stochastik und Wirtschaftsmathematik, TU Wien, https://institute.tuwien.ac.at/fileadmin/t/mathstoch/upload/mw1.pdf, zuletzt besucht: 28.2.2021

[H] Hartmann, K., Krois, J., Waske, B.: *E-Learning Project SOGA: Statistics and Geospatial Data Analysis.*, (2018), Department of Earth Sciences, Freie Universitaet Berlin, zuletzt besucht: 28.2.2021

[K] Michael Kaltenbäck: *Fundament Analysis*, (2014), Heldermann Verlag, Lemgo, Technische Universität Wien

[M] Steven J. Miller: *Benfords Law: Theory and Applications*, (2015), veröffentlicht durch: Princton University Press, https://www.researchgate.net/publication/280157559_ Benford % 27s_ Law_ Theory_ and_ Applications, zuletzt besucht: 28.2.2021

[S] Elias M. Stein, Rami Shakarchi: *Fourier Analysis: an Introduction*, (2003), Princeton University Press, New Jersey