Der Median

Klaus Pommerening

Februar 2022; letzte Änderung: 17. Februar 2022

Der Median einer endlichen Folge von reellen Zahlen ist leicht verständlich zu definieren: Man ordnet die Zahlen (einschließlich eventueller Wiederholungen) der Größe nach an und nimmt (bei ungerader Anzahl) den in der Mitte stehenden Wert bzw. (bei gerader Anzahl) den Mittelwert der beiden in der Mitte stehenden Werte.

Komplizierter, und zumindest Grundkenntnisse der reellen Analysis voraussetzend, ist die Definition des Medians einer beliebigen Wahrscheinlichkeitsverteilung auf der reellen Zahlengeraden. Sie ist der Gegenstand dieses Textes.

1 Der Median einer Verteilung

Definition Eine Funktion

$$F : \mathbb{R} \longrightarrow \mathbb{R}$$

heißt **Verteilungsfunktion**[1], wenn

(i) F monoton wächst, d. h., für $x \leq y$ ist $F(x) \leq F(y)$,

(ii) F rechtsseitig stetig ist, d. h.[2],

$$\lim_{t \searrow x} F(t) = F(x) \quad \text{für alle} \quad x \in \mathbb{R},$$

(iii) F von 0 bis 1 wächst, d. h.,

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \to \infty} F(x) = 1.$$

Insbesondere ist der Wertevorrat damit $F(\mathbb{R}) \subseteq [0, 1]$.

Definition Sei $F : \mathbb{R} \longrightarrow \mathbb{R}$ eine Verteilungsfunktion und $m \in \mathbb{R}$. Dann hat m die **Median-Eigenschaft**[3] (bezüglich F), wenn

$$\lim_{x \nearrow m} F(x) \leq \frac{1}{2} \quad \text{und} \quad F(m) \geq \frac{1}{2}. \tag{1}$$

[1] im Sinne der Wahrscheinlichkeitstheorie

[2] Der Limes von rechts (oder oben) wird hier mit dem Symbol $t \searrow x$ bezeichnet, entsprechend der Limes von links (oder unten) mit dem Symbol $t \nearrow x$.

[3] Manchmal wird m dann auch schon *ein* Median genannt. In diesem Text wird eine eindeutige Definition des Medians bevorzugt.

Die Median-Eigenschaft kann man auch so ausdrücken:

$$F(x) \le \frac{1}{2} \quad \text{für } x < m \quad \text{und} \quad F(x) \ge \frac{1}{2} \quad \text{für } x > m.$$

Diese Formulierung entspricht der intuitiven Vorstellung von einem Median: Die durch F beschriebene „Wahrscheinlichkeit“ für einen Wert $< m$ bzw. $> m$ ist jeweils höchstens 1/2. Die Asymmetrie bei der Definition der Median-Eigenschaft ist durch die Asymmetrie bei der Definition einer Verteilungsfunktion bedingt, die sich in der rechtsseitigen Stetigkeit manifestiert und auch $F(m) \ge \frac{1}{2}$ impliziert.

Definition Sei $F\colon \mathbb{R} \longrightarrow \mathbb{R}$ eine Verteilungsfunktion. Dann heißt

$$\underline{M}_F := \inf F^{-1}\left(\left[\frac{1}{2}, 1\right]\right).$$

der **Untermedian** von F.

Bemerkungen

1. $F(x) < 1/2$ für $x < \underline{M}_F$ – sonst wäre $\underline{M}_F$ nicht das Infimum. Insbesondere ist
$$\lim_{x \nearrow \underline{M}_F} F(x) \le \frac{1}{2}.$$
2. $F(x) \ge 1/2$ für $x > \underline{M}_F$.
3. Wegen der rechtsseitigen Stetigkeit ist auch $F(\underline{M}_F) \ge 1/2$, also sogar
$$\underline{M}_F = \min F^{-1}\left(\left[\frac{1}{2}, 1\right]\right),$$
und somit hat $\underline{M}_F$ die Median-Eigenschaft (1).
4. Wegen der Monotonie von F ist die Faser $F^{-1}F(\underline{M}_F)$ ein Intervall (links abgeschlossen, rechts offen oder abgeschlossen).
5. Falls F den Wert 1/2 annimmt, also nach dem Zwischenwertsatz insbesondere, wenn F stetig ist, ist
$$\underline{M}_F = \inf F^{-1}\left(\frac{1}{2}\right) = \min F^{-1}\left(\frac{1}{2}\right)$$
und somit $F(\underline{M}_F) = 1/2$. Dann hat jede Zahl $m \in F^{-1}F(\underline{M}_F)$, also mit $F(m) = F(\underline{M}_F)$, die Median-Eigenschaft (1). Von diesen Zahlen ist $\underline{M}_F$ die kleinste.
6. Falls F den Wert 1/2 nicht annimmt, ist $b := F(\underline{M}_F) > 1/2$. Egal, ob das Intervall $F^{-1}(b)$ noch weitere Elemente enthält – $\underline{M}_F$ ist das einzige davon, das die Median-Eigenschaft hat.

Die asymmetrische Definition des Untermedians wird durch ein Gegenstück etwas abgemildert:

Definition Sei $F\colon \mathbb{R} \longrightarrow \mathbb{R}$ eine Verteilungsfunktion.

(i) Die obere Intervallgrenze

$$\overline{M}_F := \sup F^{-1}F(\underline{M}_F),$$

heißt der **Obermedian** von F.

(ii) Falls F den Wert $1/2$ annimmt, heißt der Mittelwert

$$M_F := \frac{\underline{M}_F + \overline{M}_F}{2}$$

der **Median** von F.

(iii) Falls F den Wert $1/2$ nicht annimmt (also insbesondere $F(\underline{M}_F) > 1/2$), heißt $M_F := \underline{M}_F$ der **Median** von F.

Bemerkungen

1. Der Obermedian hat genau dann die Meridian-Eigenschaft, wenn er
   - mit dem Untermedian übereinstimmt oder wenn
   - $F(\overline{M}_F) = F(\underline{M}_F) = 1/2$, also wenn das Intervall $F^{-1}F(\underline{M}_F)$ abgeschlossen ist und F den Wert $1/2$ annimmt. Im Fall $\overline{M}_F > \underline{M}_F$ ist F dann an der Stelle $\overline{M}_F$ auch linksseitig stetig.
2. Der Median hat in jedem Fall die Median-Eigenschaft. Im Fall (iii) ist das trivial, im Fall (ii) folgt es, weil er im Intervall $F^{-1}(1/2)$ liegt, siehe Bemerkung 5 oben, egal, ob $\overline{M}_F =$ oder $> \underline{M}_F$.
3. Genau dann, wenn $F^{-1}F(\underline{M}_F)$ einelementig ist, gilt die Gleichheit

   $$\underline{M}_F = \overline{M}_F = M_F.$$

4. Die Definition des Medians ist etwas kompliziert, aber so passt sie zur „naiven“ Definition des Medians einer empirischen Verteilung, siehe unten.

Wir halten fest, dass der Median M_F ist für jede Verteilungsfunktion F eindeutig definiert ist und dass $F(M_F) \geq 1/2$, wobei die Gleichheit nicht notwendig gilt. Bewiesen wurde:

Satz 1 *Sei $F\colon \mathbb{R} \longrightarrow \mathbb{R}$ eine Verteilungsfunktion und M_F ihr Median. Dann gilt:*

(i) $F(M_F) = 1/2 \Longleftrightarrow F$ *nimmt den Wert* $1/2$ *an.*

(ii) *Falls F den Wert $1/2$ nicht annimmt, gibt es einen minimalen Wert $b > 1/2$, den F annimmt, $F(M_F) = b$ und $M_F = \min F^{-1}(b)$.*

Beispiel: Logistische Verteilung

Abbildung 1 zeigt als Beispiel einer Verteilungsfunktion die logistische Funktion

$$F(x) = \frac{1}{1+e^{-x}};$$

ihr Median ist 0, weil sie streng monoton wächst, insbesondere alle Fasern einelementig sind, und $F(0) = 1/2$.

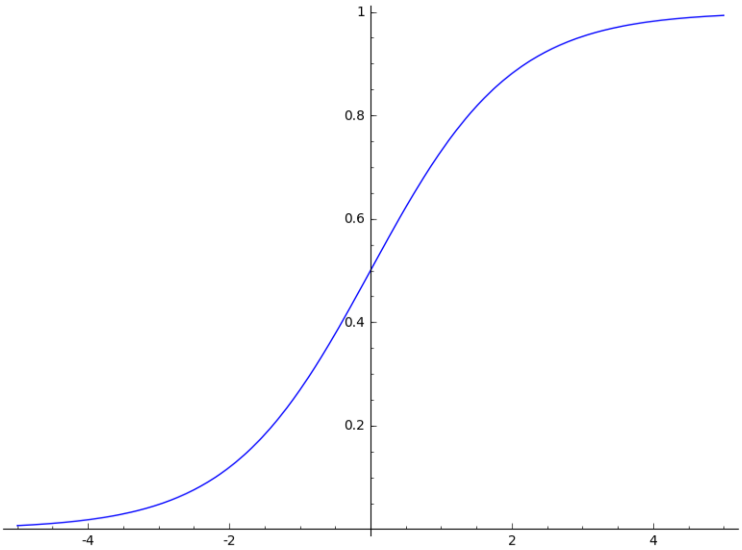

Abbildung 1: Die logistische Funktion als Verteilungsfunktion mit Median 0

2 Diskrete Verteilungen

Beispiel

Gegeben sei eine Folge[4][5]

$$(q_i)_{i\in\mathbb{N}} \quad \text{mit allen} \quad q_i \geq 0 \quad \text{und} \quad \sum_{i\in\mathbb{N}} q_i = 1.$$

Die Partialsummen seien

$$s_k := \sum_{i=0}^{k} q_i \quad \text{für } k \in \mathbb{N}.$$

[4] In diesem Text gehört 0 immer zu $\mathbb{N}$.

[5] Vorstellung: Bei zufälliger Wahl einer Zahl $\in \mathbb{N}$ soll i die Wahrscheinlichkeit q_i haben.

Damit wird eine diskrete Verteilungsfunktion definiert:

$$F(x) := s_{\lfloor x \rfloor} \quad \text{für } x \in \mathbb{R},$$

die eine (rechtsseitig stetige) Treppenfunktion ist.

Beispiel Abbildung 2 zeigt die Verteilungsfunktion für die Folge

$$q_0 = 0,\ q_1 = 0.2,\ q_2 = 0.3,\ q_3 = 0.15,\ q_4 = 0.25,\ q_5 = 0.1,\ q_i = 0 \text{ für } i \geq 6,$$

mit den Partialsummen

$$s_0 = 0,\ s_1 = 0.2,\ s_2 = 0.5,\ s_3 = 0.65,\ s_4 = 0.9,\ s_5 = 1,\ s_k = 1 \text{ für } k \geq 6.$$

Die rechtsseitige Stetigkeit wird durch die Hervorhebung der „Sprungpunkte" veranschaulicht. Insbesondere ist $F(2) = 1/2$ und $F^{-1}(1/2) = [2, 3[$. Daher ist

$$\underline{M}_F = 2, \quad \overline{M}_F = 3, \quad M_F = 2.5.$$

Eine empirische Verteilung von 20 natürlichen Zahlen, deren relative Häufigkeiten den q_i entsprächen, sähe also so aus[6]:

$$\underbrace{1, 1, 1, 1}_{4}, \underbrace{2, 2, 2, 2, 2, 2}_{6}, \underbrace{3, 3, 3}_{3}, \underbrace{4, 4, 4, 4, 4}_{5}, \underbrace{5, 5}_{2}$$

Da die Anzahl 20 der Werte gerade ist, ist der empirische Median dieser Beobachtungsreihe der Mittelwert aus der zehnten und elften Zahl, also $(2 + 3)/2 = 2.5$, in Übereinstimmung mit dem Median der Verteilungsfunktion.

Empirische endliche Verteilung

Es seien n reelle Zahlen in aufsteigender Größe,

$$x_1 < x_2 < \ldots < x_n,$$

als mögliche Werte („Beobachtungswerte") gegeben. In einer Reihe von N Zahlen („Beobachtungen") komme x_i genau N_i-mal vor, also mit relativer Häufigkeit $q_i = N_i/N$. Die Partialsummen seien wieder $s_1, \ldots, s_n$ wie oben. Die zugehörige empirische Verteilungsfunktion $F\colon \mathbb{R} \longrightarrow \mathbb{R}$ ist definiert durch

$$F(x) := \begin{cases} 0 & \text{für } x < x_1, \\ s_i & \text{für } x_i \leq x < x_{i+1} \text{ mit } 1 \leq i < n \\ 1 & \text{für } x \geq x_n. \end{cases}$$

[6] Der Wert i wird $20\, q_i$-mal angenommen.

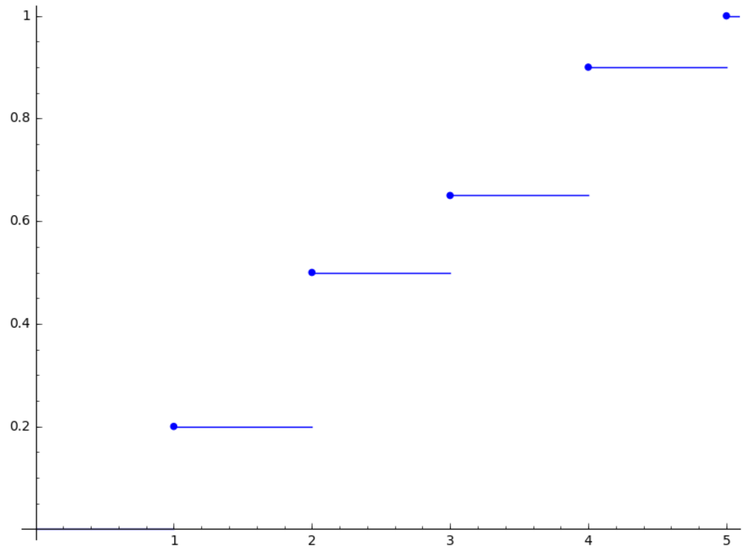

Abbildung 2: Eine diskrete Verteilungsfunktion

Zur Bestimmung des Medians sucht man den minimalen Index k mit $s_k \geq 1$. Dann ist x_k die minimale reelle Zahl x mit $F(x) \geq 1/2$. Also ist

$$\underline{M}_F = x_k, \quad F^{-1}F(\underline{M}_F) = [x_k, x_{k+1}[\quad \text{und} \quad \overline{M}_F = x_{k+1}.$$

Der Median ist also

$$M_F = \begin{cases} x_k, & \text{falls } s_k > \frac{1}{2}, \\ \frac{1}{2}\,(x_k + x_{k+1}), & \text{falls } s_k = \frac{1}{2}. \end{cases}$$

Schreibt man alle Beobachtungswerte hin:

$$\underbrace{\underbrace{x_1}_{N_1} \ldots \underbrace{x_k}_{N_k = Nq_k}}_{Ns_k} \ldots, \tag{2}$$

so sieht man, dass der Median im naiven Sinne („mittlerer Wert“ bzw. „Mittelwert der beiden mittleren Werte“) genau mit dem Median der empirischen Verteilungsfunktion übereinstimmt: Falls $s_k > 1/2$, ist der mittlere Wert (falls N ungerade) bzw. sind die beiden mittleren Werte (falls N gerade) gleich x_k. Der Fall $s_k = 1/2$ kann nur für gerades N vorkommen; dann steht x_k an der Stelle $N/2$ in der Reihe (2) und x_{k+1} an der nächsten, der („empirische“) Median ist also der Mittelwert dieser beiden und stimmt somit ebenfalls mit M_F überein.

3 Wahrscheinlichkeitsdichten

Im nichtdiskreten Fall wird eine Verteilungsfunktion meistens auf eine spezielle Weise gewonnen:

Definition Eine **Wahrscheinlichkeitsdichte** ist eine integrierbare[7] Funktion

$$f : \mathbb{R} \longrightarrow \mathbb{R}_+ = [0, \infty[\quad \text{mit} \quad \int_{\mathbb{R}} f = 1.$$

Die zugehörige Verteilungsfunktion F wird dann durch das Integral

$$F(x) := \int_{-\infty}^{x} f$$

definiert.

Beispiel Die logistische Funktion F ist differenzierbar mit Ableitung

$$f(x) := F'(x) = \frac{e^{-x}}{(1 + e^{-x})^2}.$$

Sie entsteht also durch Integration aus dieser Wahrscheinlichkeitsdichte f.

Bemerkung Für diesen Text reicht es, f als stückweise stetig anzunehmen. Dann ist f integrierbar und F stetig.

Gibt es außerdem $a, b \in \mathbb{R}$ mit $a < M_F < b$ und $f(t) > 0$ für $a < t < b$, so wächst F im Intervall $[a, b]$ streng monoton, d. h., $F^{-1}(1/2)$ ist einelementig $= \{M_F\}$. Die ganze Kompliziertheit der Definition des Medians spielt dann keine Rolle.

Hilfssatz 1 *Sei f eine Wahrscheinlichkeitsdichte und F die zugehörige Verteilungsfunktion. Es sei f symmetrisch um $a \in \mathbb{R}$, also $f(a + x) = f(a - x)$ für alle $x \in \mathbb{R}$. Dann gilt:*

(i) *$F(a + x) = 1 - F(a - x)$ für alle $x \in \mathbb{R}$.*

(ii) *a ist der Median von F.*

Beweis. (i) Durch Substitution folgt

$$\begin{aligned} F(a + x) &= \int_{-\infty}^{a+x} f(t)\,dt = \int_{-\infty}^{x} f(a + t)\,dt = \int_{-\infty}^{x} f(a - t)\,dt \\ &= \int_{-x}^{\infty} f(a + t)\,dt = \int_{-x+a}^{\infty} f(t)\,dt = 1 - \int_{-\infty}^{-x+a} f(t)\,dt = 1 - F(a - x). \end{aligned}$$

(ii) Aus (i) folgt insbesondere $F(a) = 1/2$, und wenn $F(a + x) = 1/2$ ist, so auch $F(a - x) = 1/2$. Das Intervall $F^{-1}(1/2)$ ist also symmetrisch um a und insbesondere abgeschlossen. Seine Grenzen, also der Unter- und der Obermedian, liegen daher symmetrisch zu a, und ihr Mittelwert ist $M_F = a$. $\diamond$

[7] mindestens Lebesgue-integrierbar, aber die maßtheoretischen Spitzfindigkeiten sollen hier nicht vertieft werden

Beispiel: Normal- und Lognormalverteilung

Die Standard-Normalverteilung hat die Verteilungsfunktion Φ mit Wahrscheinlichkeitsdichte

$$\varphi(t) = \Phi'(t) = \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2},$$

das ist die bekannte Gaußsche Glocke. Für Φ selbst ist die einfachste explizite Formel

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

Da φ symmetrisch um 0 ist, ist 0 der Median der Standard-Normalverteilung nach Hilfssatz 1.

Im allgemeinen Fall hängt die Normalverteilung von zwei Parametern μ und σ ab und hat die Verteilungsfunktion

$$\mathcal{N}_{\mu,\sigma}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

und die Wahrscheinlichkeitsdichte

$$\nu_{\mu,\sigma}(t) = \frac{1}{\sigma}\,\varphi\left(\frac{t-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(t-\mu)^2/2\sigma^2}.$$

Sie ist symmetrisch um μ, also ist nach Hilfssatz 1 der Median gleich μ. Wegen der Positivität von $\nu_{\mu,\sigma}$ ist $\mathcal{N}_{\mu,\sigma}^{-1}(1/2) = \{\mu\}$.

Die Lognormalverteilung [2] hat die Verteilungsfunktion

$$\mathcal{L}_{\mu,\sigma}(x) = \begin{cases} \Phi\left(\frac{\ln x-\mu}{\sigma}\right) & \text{für } x > 0, \\ 0 & \text{für } x \le 0. \end{cases}$$

Ihre Wahrscheinlichkeitsdichte ist also (für $x > 0$)

$$\mathcal{L}'_{\mu,\sigma}(x) = \varphi\left(\frac{\ln x-\mu}{\sigma}\right)\cdot\frac{1}{\sigma x} = \frac{1}{\sqrt{2\pi}\,\sigma x}\, e^{-(\ln x-\mu)^2/2\sigma^2} =: \ell_{\mu,\sigma}(x),$$

ergänzt durch $\ell_{\mu,\sigma}(x) = 0$ für $x \le 0$. Da

$$\mathcal{L}_{\mu,\sigma}(x) = \Phi\left(\frac{\ln x-\mu}{\sigma}\right) = \frac{1}{2} \Longleftrightarrow \frac{\ln x-\mu}{\sigma} = 0 \Longleftrightarrow \ln x = \mu,$$

ist e^μ der Median der Lognormalverteilung $\mathcal{L}_{\mu,\sigma}$.

Satz 2 *Der Median*

(i) *der Normalverteilung $\mathcal{N}_{\mu,\sigma}$ ist μ,*

(ii) *der Lognormalverteilung $\mathcal{L}_{\mu,\sigma}$ ist e^μ.*

Abbildung 3 zeigt die Wahrscheinlichkeitsdichte der Lognormalverteilung mit den exemplarischen Parametern $\mu = 3$ und $\sigma = 0.3$ samt dem zugehörigen Median.

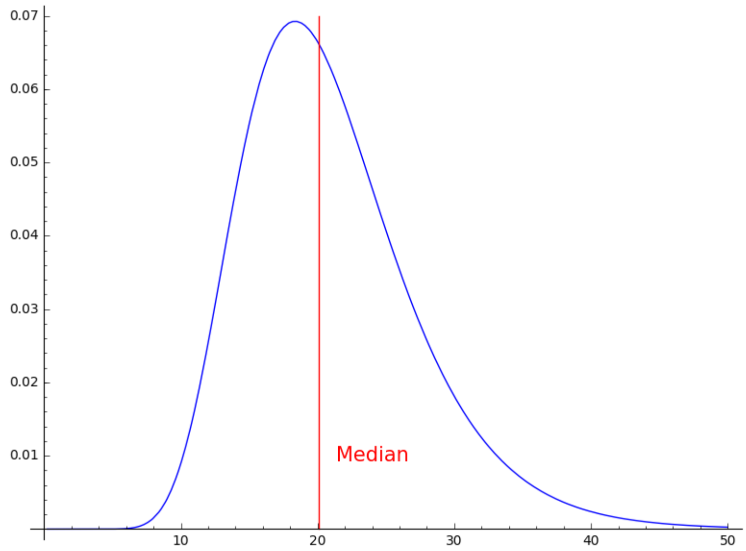

Abbildung 3: Lognormalverteilung: Wahrscheinlichkeitsdichte $\ell_{3,0.3}$

Beispiel: Gammaverteilung

Die Gammaverteilung [3] hängt von zwei Parametern $p > 0$ und $b > 0$ ab und ist definiert durch die Wahrscheinlichkeitsdichte

$$\gamma_{p,b}(x) = \begin{cases} \frac{b^p}{\Gamma(p)}\, x^{p-1}\, e^{-bx} & \text{für } x > 0, \\ 0 & \text{für } x \leq 0, \end{cases}$$

wobei Γ die Gammafunktion ist, die die Fakultät natürlicher Zahlen verallgemeinert. Die Verteilungsfunktion ist

$$\mathcal{G}_{p,b}(x) = \begin{cases} \frac{b^p}{\Gamma(p)} \int_0^x t^{p-1}\, e^{-bt}\, dt & \text{für } x > 0, \\ 0 & \text{für } x \leq 0, \end{cases} \tag{3}$$

Auch hier gibt es für die Verteilungsfunktion keine elegante explizite Formel, die eine einfache Bestimmung des Medians ermöglichen würde, und hier entfällt auch die Symmetrie-Überlegung, die diese Bestimmung für die Normalverteilung und als Folge davon auch für die Lognormalverteiung trivial gemacht hat. Als Methode zur Bestimmung des Medians bleibt also bis auf weiteres nur, die Gleichung

$$\int_0^x t^{p-1}\, e^{-bt} = \frac{\Gamma(p)}{2b^p}$$

nach x aufzulösen. Substituiert man in (3) die Integrationsvariable $t = s/b$, so sieht man, dass

$$\mathcal{G}_{p,b}(x) = \mathcal{G}_{p,1}(bx).$$

Hilfssatz 2 *Bezeichnet man den Median von $\mathcal{G}_{p,b}$ mit $M_{p,b}$, so gilt:*

$$M_{p,b} = \frac{1}{b}\, M_{p,1},$$

Beweis. Das ist genau die Stelle, an der $\mathcal{G}_{p,b}$ den Wert $1/2$ annimmt. $\diamond$

Durch diesen Hilfssatz ist das Problem der Medianbestimmung auf den Fall $b = 1$ reduziert, also auf die Lösung $x = M_{p,1}$ von

$$\int_0^x t^{p-1}\, e^{-t} = \frac{\Gamma(p)}{2}.$$

Die linke Seite ist die bekannte unvollständige Gamma-Funktion $\Gamma(p, x)$.

Satz 3 *Der Median der Gammaverteilung $\mathcal{G}_{p,b}$ ist x/p, wobei x die Lösung der Gleichung*

$$\Gamma(p, x) = \frac{\Gamma(p)}{2}$$

ist.

Abbildung 4 zeigt die Wahrscheinlichkeitsdichte der Gammaverteilung mit den exemplarischen Parametern $p = 2$ und $b = 0.1$ samt dem zugehörigen Median.

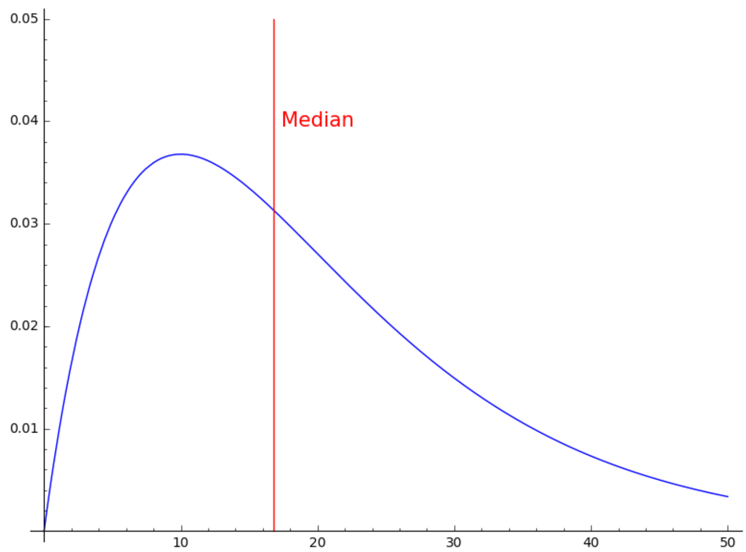

Abbildung 4: Gammaverteilung: Wahrscheinlichkeitsdichte $\gamma_{2,0.1}$

Etwas anschaulichere Ergebnisse stammen von Choi [1] für ganzzahliges p: enge obere und untere Schranken sowie eine asymptotische Entwicklung in p. Eines der Ergebnisse besagt insbesondere, dass für ganzzahliges $p \geq 19$

$$p - \frac{1}{3} < M_{p,1} \leq p - \frac{1}{3} + \frac{1}{p}.$$

Beispiel: Rayleigh-Verteilung

Die Rayleigh-Verteilung [4] hängt von einem Parameter $\sigma > 0$ ab und ist durch die Verteilungsfunktion

$$\mathcal{R}_\sigma(x) = \begin{cases} 1 - e^{-x^2/2\sigma^2} & \text{für } x > 0, \\ 0 & \text{für } x \leq 0, \end{cases}$$

definiert. Ihre Wahrscheinlichkeitsdichte ist

$$\rho_\sigma(x) = \mathcal{R}'_\sigma(x) = \begin{cases} \frac{x}{\sigma^2}\, e^{-x^2/2\sigma^2} & \text{für } x > 0, \\ 0 & \text{für } x \leq 0. \end{cases}$$

Da diese für $x > 0$ durchweg positiv ist, ist zur Bestimmung des Medians nur die Gleichung $\mathcal{R}_\sigma(x) = 1/2$ nach x aufzulösen:

$$\mathcal{R}_\sigma(x) = \frac{1}{2} \iff e^{-\frac{x^2}{2\sigma^2}} = \frac{1}{2} \iff \frac{x^2}{2\sigma^2} = \ln 2 \iff \frac{x^2}{\sigma^2} = \ln 4 \iff x = \sigma\sqrt{\ln 4}.$$

Satz 4 *Der Median der Rayleigh-Verteilung $\mathcal{R}_\sigma$ ist $\sigma\sqrt{\ln 4}$.*

Abbildung 5 zeigt die Wahrscheinlichkeitsdichte der Rayleigh-Verteilung mit dem exemplarischen Parameter $\sigma = 11$ samt dem zugehörigen Median.

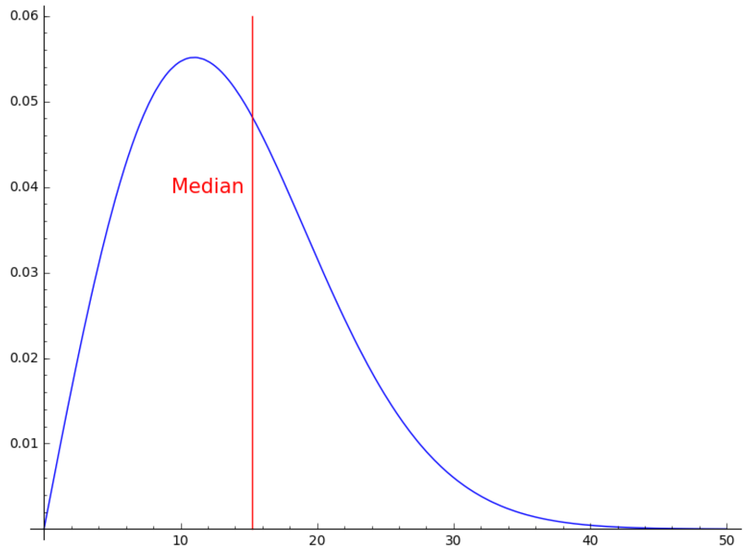

Abbildung 5: Rayleigh-Verteilung: Wahrscheinlichkeitsdichte ρ_{11}

Beispiel: Pareto-Verteilung

Die Pareto-Verteilung [5] hängt von zwei Parametern $m > 0$ und $k > 0$ ab und ist durch die Verteilungsfunktion

$$\mathcal{P}_{m,k}(x) = \begin{cases} 1 - \frac{m^k}{x^k} & \text{für } x \geq m, \\ 0 & \text{für } x < m, \end{cases}$$

definiert. Die Wahrscheinlichkeitsdichte ist

$$\wp_{m,k}(x) = \begin{cases} \frac{k\, m^k}{x^{k+1}} & \text{für } x \geq m, \\ 0 & \text{für } x < m. \end{cases}$$

Diese ist für $x \geq m$ durchweg positiv, der Median wird also wieder durch Gleichungsauflösung bestimmt:

$$\mathcal{P}_{m,k}(x) = \frac{1}{2} \Longleftrightarrow x^k = 2\, m^k \Longleftrightarrow x = m\, 2^{1/k}.$$

Satz 5 *Der Median der Pareto-Verteilung $\mathcal{P}_{m,k}$ ist $m\, 2^{1/k}$.*

Abbildung 6 zeigt die Wahrscheinlichkeitsdichte der Pareto-Verteilung mit den exemplarischen Parametern $m = 10$ und $k = 0.5$ samt dem zugehörigen Median $10 \cdot 2^2$. Der Median wirkt auf den ersten Blick ziemlich groß (d. h., nach rechts gerückt). Das liegt daran, dass die Pareto-Verteilung endlastig ist (‘fat tail’) – besonders für kleines k; die Wahrscheinlichkeitsdichte nimmt wesentlich schwächer ab, nämlich reziprok-polynomial, als bei den Verteilungen, die reziprok-exponentiell abnehmen.

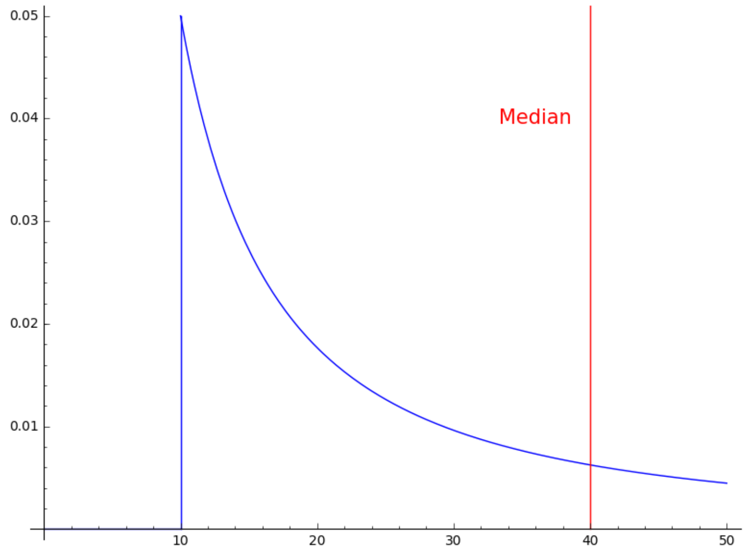

Abbildung 6: Pareto-Verteilung: Wahrscheinlichkeitsdichte $\wp_{10,0.5}$

Literatur

[1] K. P. Choi: On the medians of Gamma distributions and an equation of Ramanujan. Proc. Amer. Math. Soc. 121 (1994), 245–251

[2] Wikipedia: Logarithmische Normalverteilung. `https://de.wikipedia.org/wiki/Logarithmische_Normalverteilung` [aufgerufen am 16. Januar 2022]

[3] Wikipedia: Gammaverteilung. `https://de.wikipedia.org/wiki/Gammaverteilung` [aufgerufen am 1. November 2021]

[4] Wikipedia: Rayleigh-Verteilung. `https://de.wikipedia.org/wiki/Rayleigh-Verteilung` [aufgerufen am 1. November 2021]

[5] Wikipedia: Pareto-Verteilung. `https://de.wikipedia.org/wiki/Pareto-Verteilung` [aufgerufen am 1. November 2021]