Statistik für Informatiker, SS 2018

1.4 Erwartungswert, Varianz und Kovarianz

1.4.4 Median(e)

Matthias Birkner

http://www.staff.uni-mainz.de/birkner/StatInfo18/

4.6.2018

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Anschaulich ist der Median einer reellen Zufallsvariable X der Wert m, so dass

$$„P(X \leq m) = \frac{1}{2} = P(X \geq m)“$$

gilt.

Anschaulich ist der Median einer reellen Zufallsvariable X der Wert m, so dass

$$„P(X \le m) = \frac{1}{2} = P(X \ge m)“$$

gilt.

Da man diese Gleichheit (zumal im diskreten Fall) nicht immer genau einstellen kann, definiert man formal folgendermaßen:

Definition 1.85

X reelle ZV, m heißt (ein) Median von X (auch „Zentralwert“, manchmal auch m_X geschrieben), wenn gilt

$$P(X \ge m) \ge \frac{1}{2} \quad \text{und} \quad P(X \le m) \ge \frac{1}{2}.$$

m Median von X, wenn

$$P(X \geq m) \geq \frac{1}{2} \quad \text{und} \quad P(X \leq m) \geq \frac{1}{2}.$$

Ein Median existiert stets, auch wenn X keinen Erwartungswert besitzt.

m Median von X, wenn

$$P(X \geq m) \geq \frac{1}{2} \quad \text{und} \quad P(X \leq m) \geq \frac{1}{2}.$$

Ein Median existiert stets, auch wenn X keinen Erwartungswert besitzt.

Man kann den Median als eine „robustere“ Antwort auf die Aufgabe, für eine ZV *nur einen* „typischen Wert“ anzugeben, ansehen (im Gegensatz zum Erwartungswert besitzt ja jede Verteilung einen Median).

m Median von X, wenn

$$P(X \geq m) \geq \frac{1}{2} \quad \text{und} \quad P(X \leq m) \geq \frac{1}{2}.$$

Ein Median existiert stets, auch wenn X keinen Erwartungswert besitzt.

Man kann den Median als eine „robustere“ Antwort auf die Aufgabe, für eine ZV *nur einen* „typischen Wert“ anzugeben, ansehen (im Gegensatz zum Erwartungswert besitzt ja jede Verteilung einen Median).

Allerdings gibt es für Mediane keine so angenehmen Rechenregeln, wie sie Satz 1.70 für den Erwartungswert liefert.

m Median von X, wenn

$$P(X \geq m) \geq \frac{1}{2} \quad \text{und} \quad P(X \leq m) \geq \frac{1}{2}.$$

Bemerkung

Falls der Median von X (im Sinne von Def. 1.85) uneindeutig ist, d.h. wenn die Verteilungsfunktion F_X von X den Wert $\frac{1}{2}$ auf einem nicht-trivialen Intervall annimmt, so betrachtet man gelegentlich „pragmatisch“

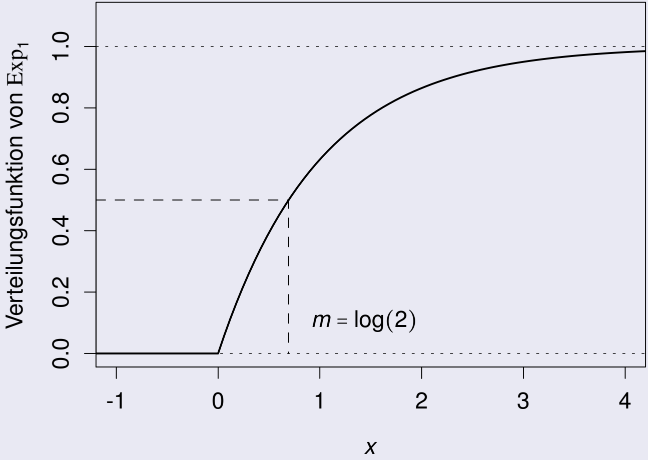
$$\frac{1}{2}\Big(\inf\{x \in \mathbb{R} : F_X(x) = \tfrac{1}{2}\} + \sup\{x \in \mathbb{R} : F_X(x) = \tfrac{1}{2}\}\Big),$$

das arithmetische Mittel des kleinst- und des größtmöglichen Medians, als „den“ Median.

(So berechnet es beispielsweise `R`.)

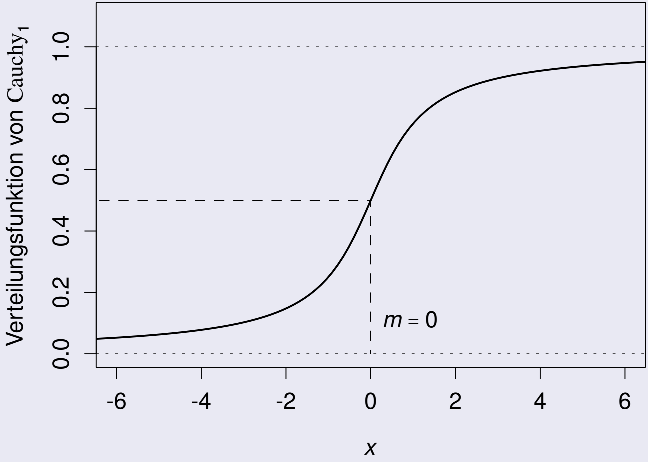
Beispiel 1.86

1. $X \sim \mathrm{Exp}_\theta$ hat Dichte $\theta e^{-\theta x}\mathbf{1}_{[0,\infty)}(x)$, Verteilungsfunktion $(1 - e^{-\theta x})\mathbf{1}_{[0,\infty)}(x)$, demnach ist der (eindeutig bestimmte) Median $m = \frac{1}{\theta}\log 2$.



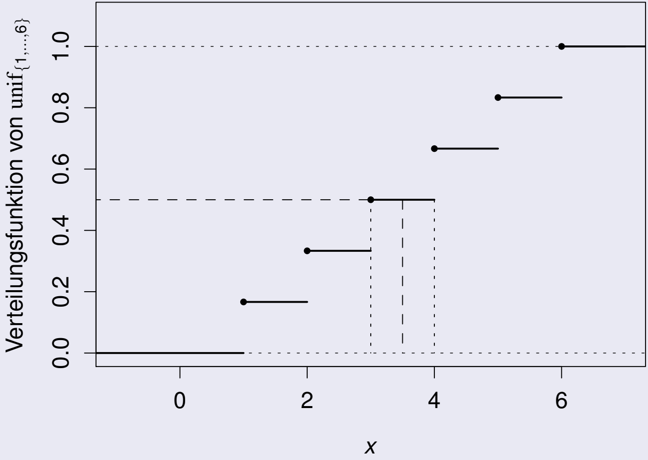
Beispiel 1.86 (Fortsetzung)

2. X Cauchy-verteilt mit Dichte $\frac{1}{\pi}\frac{1}{1+x^2}$, Verteilungsfunktion $\frac{1}{2}\frac{1}{\pi}\arctan(x)$, der (eindeutig bestimmte) Median ist $m = 0$.



Beispiel 1.86 (Fortsetzung)

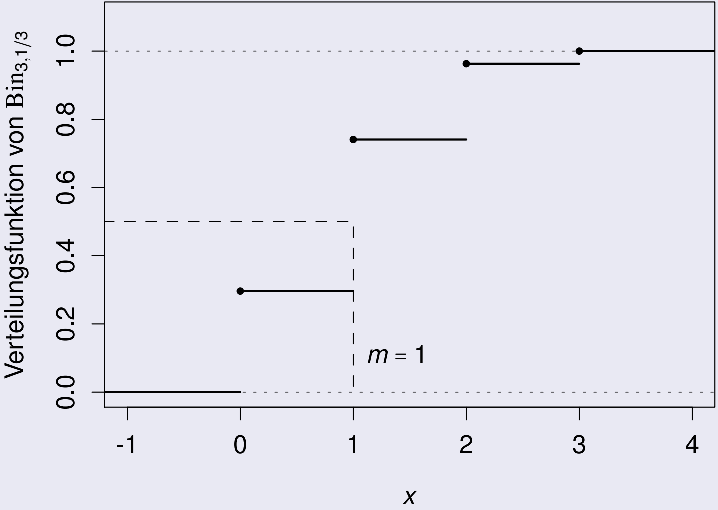
3. $X \sim \mathrm{unif}_{\{1,2,\ldots,6\}}$



Jeder Wert $m \in [3,4]$ ist ein Median (und die vielleicht „kanonischste“ Wahl wäre $m = 3{,}5$).

Beispiel 1.86 (Fortsetzung)

4. $X \sim \mathrm{Bin}_{3,1/3}$ hat Median 1



Bemerkung 1.87

Sei $X \in \mathcal{L}^1$.

1. Jeder Median von X ist ein Minimierer von
   $a \mapsto \mathbb{E}\big[|X - a|\big]$.
2. Für jeden Median m ist $\big|\mathbb{E}[X] - m\big| \leq \sqrt{\mathrm{Var}[X]}$.

Bemerkung 1.87

Sei $X \in \mathcal{L}^1$.

1. Jeder Median von X ist ein Minimierer von $a \mapsto \mathbb{E}\big[|X-a|\big]$.
2. Für jeden Median m ist $\big|\mathbb{E}[X] - m\big| \le \sqrt{\mathrm{Var}[X]}$.

Beweis. 1. Sei m ein Median. Falls $a > m$:

$$|X-a| - |X-m| \ge (a-m)\mathbf{1}_{\{X \le m\}} - (a-m)\mathbf{1}_{\{X > m\}},$$

also

$$\mathbb{E}\big[|X-a|\big] - \mathbb{E}\big[|X-m|\big] \ge (a-m)\Big(\underbrace{P(X \le m)}_{\ge 1/2} - \underbrace{P(X > m)}_{\le 1/2}\Big) \ge 0,$$

analog im Fall $a < m$.

Bemerkung 1.87

Sei $X \in \mathcal{L}^1$.

1. Jeder Median von X ist ein Minimierer von

   $$a \mapsto \mathbb{E}\big[|X - a|\big].$$

2. Für jeden Median m ist $\big|\mathbb{E}[X] - m\big| \leq \sqrt{\mathrm{Var}[X]}$.

Bemerkung 1.87

Sei $X \in \mathcal{L}^1$.

1. Jeder Median von X ist ein Minimierer von
   $a \mapsto \mathbb{E}\big[|X - a|\big]$.

2. Für jeden Median m ist $\big|\mathbb{E}[X] - m\big| \leq \sqrt{\mathrm{Var}[X]}$.

Beweis. 2. Es ist

$$\big|\mathbb{E}[X] - m\big| = \big|\mathbb{E}[X - m]\big| \leq \mathbb{E}\big[|X - m|\big]$$

$$\overset{1.}{\leq} \mathbb{E}\big[|X - \mathbb{E}[X]|\big] = \sqrt{\Big(\mathbb{E}\big[|X - \mathbb{E}[X]|\big]\Big)^2} \leq \sqrt{\mathbb{E}\big[|X - \mathbb{E}[X]|^2\big]},$$

wobei für die erste Ungleichung die Monotonie des Erwartungswerts (beachte: $X - m \leq |X - m|$ und $-(X - m) \leq |X - m|$) und für die letzte Ungleichung die Cauchy-Schwarz-Ungleichung verwenden.