Kontinuierlich uniform verteilte Zufallsvariable

Ergänzung zu Vorlesung 2a, 19.04.2011

kontinuierlich uniform verteilte Zufallsvariable

kontinuierlich uniform verteilte Zufallsvariable

Beispiel Dartspiel: Treffpunkt sei rein zufällig in der Fläche A.

Definition: kontinuierlich uniforme Verteilung

Definition: kontinuierlich uniforme Verteilung

$S \subset \mathbb{R}^d$ mit endlichem Inhalt $V(S)$. Eine S-wertige ZV X heißt *uniform verteilt*, wenn für alle $A \subset S$ mit (wohldefiniertem) Inhalt $V(A)$ gilt

Definition: kontinuierlich uniforme Verteilung

$S \subset \mathbb{R}^d$ mit endlichem Inhalt $V(S)$. Eine S-wertige ZV X heißt *uniform verteilt*, wenn für alle $A \subset S$ mit (wohldefiniertem) Inhalt $V(A)$ gilt

$$\mathbf{P}(X \in A) = \frac{V(A)}{V(S)}.$$

Definition: kontinuierlich uniforme Verteilung

$S \subset \mathbb{R}^d$ mit endlichem Inhalt $V(S)$. Eine S-wertige ZV X heißt *uniform verteilt*, wenn für alle $A \subset S$ mit (wohldefiniertem) Inhalt $V(A)$ gilt

$$\mathbf{P}(X \in A) = \frac{V(A)}{V(S)}.$$

Man schreibt auch:

$$\mathbf{P}(X \in da) = \frac{da}{V(S)} \qquad , \qquad a \in S.$$

Beispiel: zufällige Zahl im Intervall $[a, b)$, $a, b \in \mathbb{R}$, $a < b$

Beispiel: zufällige Zahl im Intervall $[a, b)$, $a, b \in \mathbb{R}$, $a < b$

Dann gilt für $a < c < d < b$:

$$\mathbf{P}(U \in (c, d]) = \frac{d - c}{b - a}. \tag{1}$$

Benford's Gesetz

Benford's Gesetz

Frage: Wie ist die Anfangsziffer in der Dezimaldarstellung einer zufälligen Zahl X (die auf logarithmischer Skala 'breit streut') verteilt?

Benford's Gesetz

Frage: Wie ist die Anfangsziffer in der Dezimaldarstellung einer zufälligen Zahl X (die auf logarithmischer Skala 'breit streut') verteilt?

Uniform?

NEIN!

Formulierung des Problems

Sei $a > 0$ feste Zahl und $h(a)$ Anfangsziffer ihrer Dezimaldarstellung:

Formulierung des Problems

Sei $a > 0$ feste Zahl und $h(a)$ Anfangsziffer ihrer Dezimaldarstellung:

$h(a) = b$, $b = 1, \dots, 9$ genau dann, wenn

Formulierung des Problems

Sei $a > 0$ feste Zahl und $h(a)$ Anfangsziffer ihrer Dezimaldarstellung:

$h(a) = b$, $b = 1, \dots, 9$ genau dann, wenn
$b \cdot 10^n \leq a < (b + 1) \cdot 10^n$ für ein $n \in \mathbb{Z}$, bzw.

Formulierung des Problems

Sei $a > 0$ feste Zahl und $h(a)$ Anfangsziffer ihrer Dezimaldarstellung:

$h(a) = b$, $b = 1, \dots, 9$ genau dann, wenn
$b \cdot 10^n \leq a < (b+1) \cdot 10^n$ für ein $n \in \mathbb{Z}$, bzw.

$$\log_{10} b + n \leq \log_{10} a < \log_{10}(b+1) + n.$$

Formulierung des Problems

Sei $a > 0$ feste Zahl und $h(a)$ Anfangsziffer ihrer Dezimaldarstellung:

$h(a) = b$, $b = 1, \dots, 9$ genau dann, wenn
$b \cdot 10^n \leq a < (b+1) \cdot 10^n$ für ein $n \in \mathbb{Z}$, bzw.

$$\log_{10} b + n \leq \log_{10} a < \log_{10}(b+1) + n.$$

Da $\log_{10} b \in [0, 1]$ ist $n = \lfloor \log_{10} a \rfloor$ und

$$\log_{10} b \leq \log_{10} a - \lfloor \log_{10} a \rfloor < \log_{10}(b+1).$$

Annahme: $U := \log_{10} X - \lfloor \log_{10} U \rfloor$ ist uniform (auf $[0, 1)$)

Annahme: $U := \log_{10} X - \lfloor \log_{10} U \rfloor$ ist uniform (auf $[0,1)$)

Also folgt für $b = 1, \ldots, 9$,

$$\{h(X) = b\} = \left\{ U \in \left[\log_{10} b, \log_{10}(b+1) \right) \right\}$$

Annahme: $U := \log_{10} X - \lfloor \log_{10} U \rfloor$ ist uniform (auf $[0,1)$)

Also folgt für $b = 1, \dots, 9$,

$$\{h(X) = b\} = \{U \in [\log_{10} b, \log_{10}(b+1))\}$$

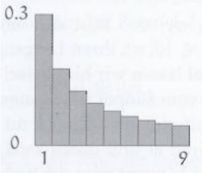
und

$$\mathbf{P}(h(X) = b) = \log_{10}(b+1) - \log_{10} b = \log_{10}\left(1 + \frac{1}{b}\right).$$

$$\mathbb{P}(h(X) = 1) \approx 0,30 \text{ und } \mathbb{P}(h(X) = 9) \approx 0,05$$

Benfords Gesetz

$\mathbb{P}(h(X) = 1) \approx 0,30$ und $\mathbb{P}(h(X) = 9) \approx 0,05$



Anwendungen

z.B.: Aufdeckung von Betrügereien bei Krankenkassen
Indizien für z.B. Wahlbetrug

Anwendungen

z.B.: Aufdeckung von Betrügereien bei Krankenkassen
Indizien für z.B. Wahlbetrug

Einschränkung:

Nur für Daten, für die es KEINE ausgezeichnete Skala gibt.
ungeeignet für z.B. Preise (an Währung adjustiert)