⟨StatWth17⟩

3. Stochastische Größen und Verteilungen

Werner Gurker

Institut für Stochastik und Wirtschaftsmathematik

Technische Universität Wien

Stochastische Größe

Wahrscheinlichkeitsraum: $(\Omega, \mathcal{A}, P)$

Eine Abbildung X von Ω nach $\mathbb{R}$:

$$X : \Omega \longrightarrow \mathbb{R}$$

die jedem $\omega \in \Omega$ eine reelle Zahl $X(\omega) = x$ zuordnet:

$$\omega \in \Omega \longmapsto X(\omega) = x \in \mathbb{R}$$

nennt man eine stochastische Größe (kurz sG).

Die Zahl $x \in \mathbb{R}$ nennt man die Realisation von X.

Merkmalraum einer sG X

Den Wertebereich von X nennt man den Merkmalraum der sG:

$$M_X = \{x \,|\, x = X(\omega),\ \omega \in \Omega\}$$

Ist M_X endlich oder abzählbar unendlich, ist die sG X diskret.

Ist M_X ein endliches oder unendliches Intervall, ist die sG X stetig (oder kontinuierlich).

In der Praxis spielen auch Mischtypen eine Rolle.

Messbarkeit

Eine Abbildung $X : \Omega \longrightarrow \mathbb{R}$ ist messbar, wenn das Urbild jeder Borelmenge $B \in \mathcal{B}$ ein Element von $\mathcal{A}$ ist:

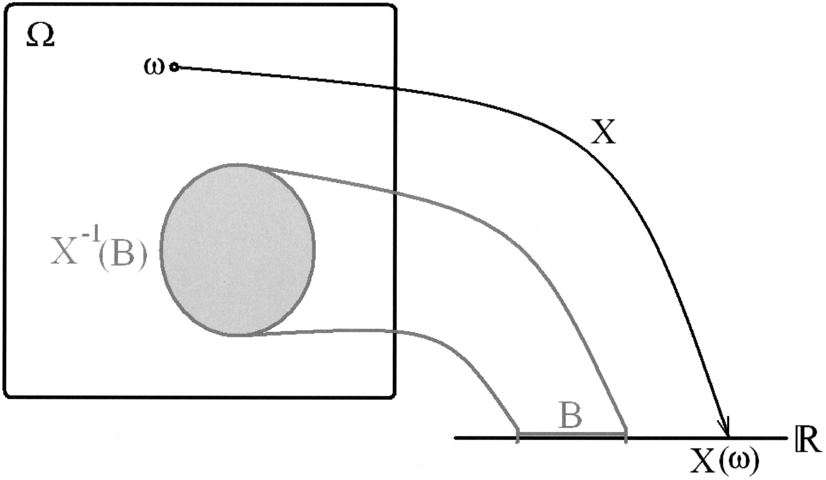
$$X^{-1}(B) = \left\{\omega \in \Omega \,\middle|\, X(\omega) \in B\right\} \in \mathcal{A} \quad \text{für alle} \quad B \in \mathcal{B}$$

Festlegung: SGn sind immer messbar.

Um die Messbarkeit von X zum Ausdruck zu bringen, schreibt man:

$$X : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B})$$

Bildliche Darstellung einer sG X



Verteilung einer sG X

A sei ein Ereignis in $\mathbb{R}$, d. h. $A \in \mathcal{B}$.

Auf Grund der Messbarkeit von X gilt:

$$X^{-1}(A) = \{\omega \in \Omega \,|\, X(\omega) \in A\} \in \mathcal{A}$$

Für $X^{-1}(A)$ schreibt man auch $X \in A$ und definiert:

$$P_X(A) := P(X \in A) = P\big(X^{-1}(A)\big)$$

P_X nennt man die – durch P *induzierte* – Verteilung von X.

Beispiel

Zwei Würfel – einer ist der “erste” und der andere ist der “zweite” – werden geworfen.

Merkmalraum: $\Omega = \{(i,j) \,|\, i,j = 1,2,\ldots,6\}$

σ–Algebra: $\mathcal{A} = \mathcal{P}(\Omega)$ (Potenzmenge)

W–Maß: $P(\{(i,j)\}) = \dfrac{1}{36}$ (Laplaceraum)

Die sG X sei die Augensumme:

$$X : (i,j) \longmapsto \; i+j$$

Beispiel (Forts.)

X ist – trivialerweise – messbar.

Das Ereignis $X = x$ ist gegeben durch:

$$\{X = x\} = \{(i,j) \,|\, i + j = x\} = X^{-1}(\{x\})$$

$$\text{z. B.:} \quad \{X = 8\} = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$$

$$\Longrightarrow \quad P_X(X = 8) = P\big(X^{-1}(\{8\})\big) = \frac{5}{36}$$

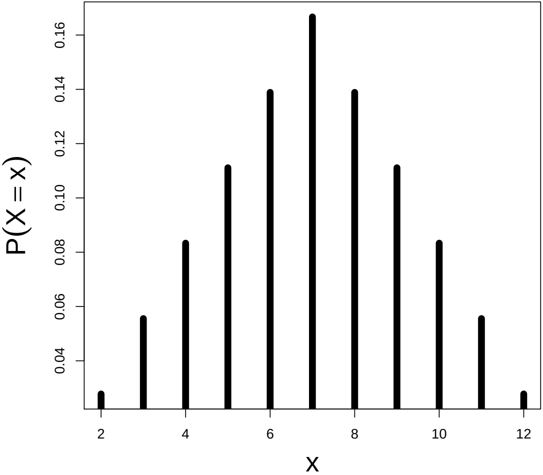
$$\text{allgemein:} \quad P(X = x) = \frac{6 - |7 - x|}{36}, \quad x = 2, 3, \ldots, 12$$

Beispiel (Forts.)

$$P(X = x) =$$

$$= \frac{6 - |7 - x|}{36}$$

$$x = 2, 3, \ldots, 12$$



Punkt– und Intervallereignisse

Wichtig in Anwendungen sind speziell:

Punktereignisse: $\{x\}, \quad x \in \mathbb{R}$

Intervallereignisse: $A = (a, b], \quad a < b$

$$P_X\big((a,b]\big) = P\big(X \in (a,b]\big) = P(a < X \leq b)$$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$

$$P_X\big(\{x\}\big) = P(X = x) = P(X \leq x) - P(X < x)$$

Verteilungsfunktion

Die Verteilungsfunktion (kurz VF) einer sG X ist definiert durch:

$$F_X(x) := P(X \leq x) \quad \text{für alle} \quad x \in \mathbb{R}$$

Eine VF F hat die folgenden Eigenschaften:

(1) $0 \leq F(x) \leq 1$ für $x \in \mathbb{R}$

(2) F ist monoton wachsend, d. h. aus $x < y$ folgt $F(x) \leq F(y)$

(3) $\lim\limits_{x \to -\infty} F(x) = 0$ und $\lim\limits_{x \to \infty} F(x) = 1$

(4) F ist rechtsstetig, d. h. $\lim\limits_{h \downarrow 0} F(x+h) = F(x)$ für $x \in \mathbb{R}$

Behauptung 1

Ist X eine sG mit Verteilungsfunktion F_X, so gilt:

$$P(a < X \leq b) = F_X(b) - F_X(a) \quad \text{für} \quad a < b$$

Beweis:

$$\{-\infty < X \leq b\} = \underbrace{\{-\infty < X \leq a\} \cup \{a < X \leq b\}}_{\text{disjunkt}}$$

$$\Longrightarrow \quad \underbrace{P(X \leq b)}_{F_X(b)} = \underbrace{P(X \leq a)}_{F_X(a)} + P(a < X \leq b)$$

Behauptung 2

Mit $F(x-) := \lim_{h \downarrow 0} F(x-h)$ (= linksseitiger Grenzwert) gilt:

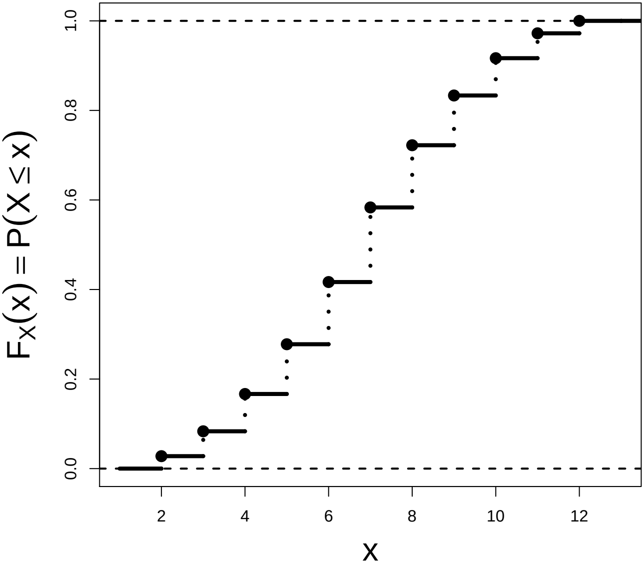
$$P(X = x) = F(x) - F(x-) \quad \text{für} \quad x \in \mathbb{R}$$

Beweis: Skriptum (S. 115)

Aus Behauptung 2 folgt:

$$F \text{ stetig an der Stelle } x \quad \Longrightarrow \quad P(X = x) = 0$$

Beispiel: Augensumme zweier Würfel



Verallgemeinerte Inverse

Häufig benötigt man die Umkehrfunktion einer VF F.

Eine VF ist aber nicht notwendig streng monoton wachsend.

Man benötigt die verallgemeinerte Inverse von F:

$$F^{-1}(y) := \inf \left\{ x \,|\, F(x) \geq y \right\} \quad \text{für} \quad y \in (0,1)$$

Ist F streng monoton wachsend, stimmt F^{-1} mit der (üblichen) Umkehrfunktion von F überein.

Quantilenfunktion

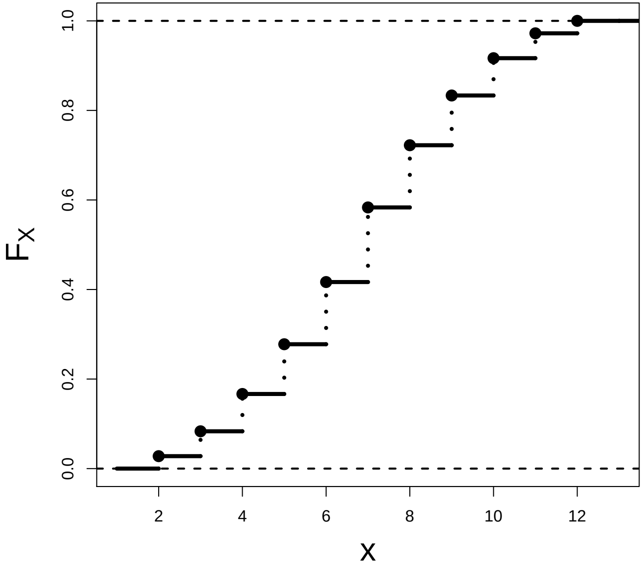
Die Funktion, die jedem $p \in (0,1)$ den Wert $F^{-1}(p)$ zuordnet, nennt man die Quantilenfunktion:

$$x_p = F^{-1}(p) \quad \text{für} \quad p \in (0,1)$$

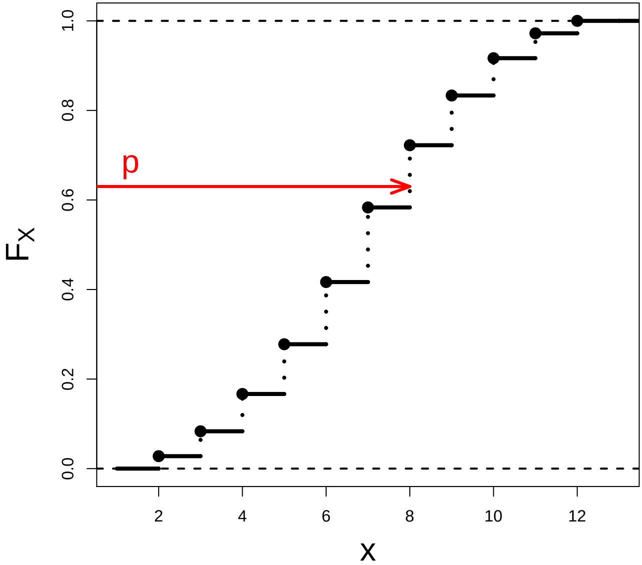
$$F^{-1} \; \ldots \text{ verallgemeinerte Inverse von } F$$

$$x_p \; \ldots \; p\text{–Quantil von } F \text{ oder } X \text{ (wenn } F = F_X)$$

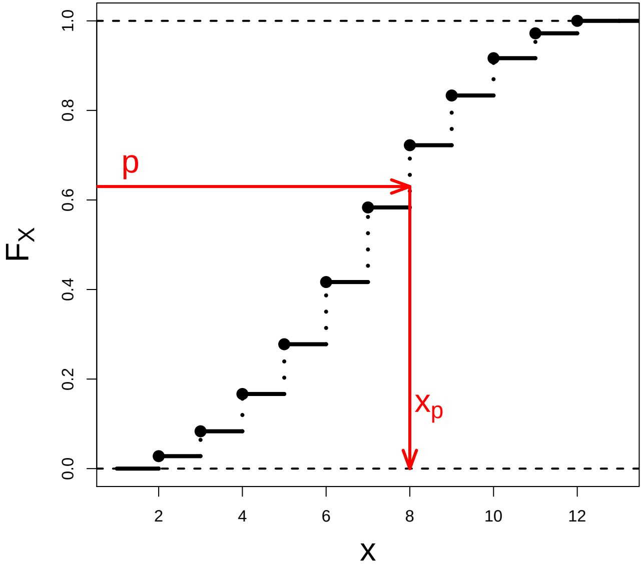
Beispiel: Augensumme zweier Würfel



Beispiel: Augensumme zweier Würfel



Beispiel: Augensumme zweier Würfel



Diskrete Verteilung

Eine sG X hat eine diskrete Verteilung (oder ist diskret), wenn ihr Merkmalraum eine endliche oder abzählbare Menge ist:

$$M_X = \{x_1, x_2, \dots\}$$

Beispiele:

Augensumme beim Werfen zweier Würfel

Werfen einer Münze bis zum ersten Auftreten von “Kopf”

Zahl der Lackierungsfehler auf einem Autoblech

etc.

Wahrscheinlichkeitsfunktion

Ist X eine diskrete sG, nennt man die Funktion, die jedem $x \in M_X$ die zugehörige Wahrscheinlichkeit zuordnet:

$$p_X(x) = P(X = x) \quad \text{für} \quad x \in M_X$$

die Wahrscheinlichkeitsfunktion (kurz W–Funktion) von X.

Bsp: X = Augensumme beim Werfen zweier Würfel

$$p_X(x) = \frac{6 - |7 - x|}{36} \quad \text{für} \quad x = 2, 3, \ldots, 12$$

Wahrscheinlichkeitsfunktion (Forts.)

Eigenschaften einer W–Funktion:

(1) $0 \leq p_X(x) \leq 1, \; x \in M_X$

(2) $\sum_{x \in M_X} p_X(x) = 1$

Wahrscheinlichkeit für ein Ereignis $B \subseteq M_X$:

$$P(X \in B) = \sum_{x \in B} p_X(x)$$

Verteilungsfunktion

X eine diskrete sG mit Merkmalraum $M_X = \{x_1, x_2, \ldots\}$.

Die VF von X ist eine Treppenfunktion:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p_X(x_i) \quad \text{für} \quad x \in \mathbb{R}$$

F_X hat Sprünge an den Stellen $x \in M_X$.

Zwischen den Sprungstellen ist F_X konstant.

Wegen der Rechtsstetigkeit von F_X gilt bei Sprungstellen immer der obere Punkt.

Stetige Verteilung

Eine sG X hat eine stetige Verteilung (oder ist stetig), wenn ihre VF $F_X(x)$ eine stetige Funktion auf $\mathbb{R}$ ist.

Punktförmige Ereignisse haben die Wahrscheinlichkeit Null:

$$\text{Behauptung 2} \quad \Longrightarrow \quad P(X = x) = F_X(x) - \underbrace{F_X(x-)}_{= F_X(x)} = 0$$

Intervalle $\langle a, b \rangle$ (mit $a < b$) haben positive Wahrscheinlichkeiten:

$$P\big(X \in \langle a, b \rangle\big) \geq 0$$

Absolute Stetigkeit

Wir nehmen an, dass stetige sGn absolut stetig sind.

D. h., es gibt eine Funktion f_X, sodass:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt$$

$$\Longrightarrow \quad \frac{dF_X(x)}{dx} = F_X'(x) = f_X(x)$$

Eine Funktion f_X mit dieser Eigenschaft nennt man Dichtefunktion (kurz Dichte) von X.

Dichtefunktion

Eine Dichte f_X hat die folgenden Eigenschaften:

(1) f_X ist nichtnegativ:

$$f_X(x) \geq 0 \quad \text{für} \quad x \in \mathbb{R}$$

(2) Die Fläche unter f_X ist Eins:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Dichtefunktion (Forts.)

Wahrscheinlichkeit für ein Ereignis $B \in \mathcal{B}$:

$$P(X \in B) = \int_B f_X(x)\, dx$$

Für ein (beliebiges) Intervall $B = \langle a, b\rangle$ $(a < b)$ gilt:

$$P\big(X \in \langle a, b\rangle\big) = F_X(b) - F_X(a) = \int_a^b f_X(t)\, dt$$

Da für stetige sGn einzelne Punkte die W. Null haben, spielt es keine Rolle ob die Randpunkte a, b des Intervalls zu B gehören oder nicht.

Träger

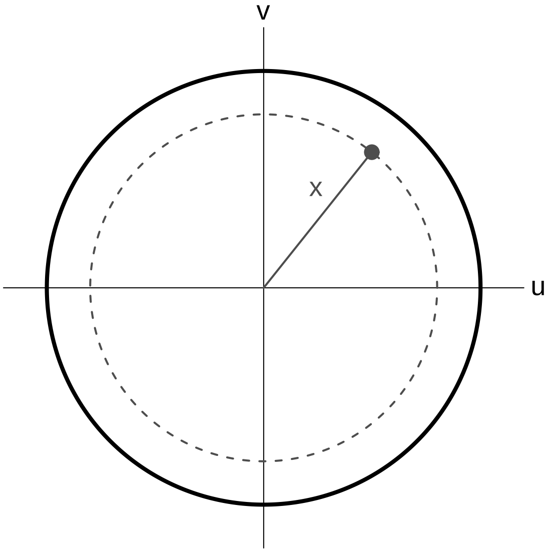
Die Menge $\mathcal{S}_X$ aller Punkte $x \in \mathbb{R}$ mit positiver Dichte nennt man den Träger von X:

$$\mathcal{S}_X = \{x : f_X(x) > 0\}$$

Der Träger ist meist ein (endliches oder unendliches) Intervall:

$$\langle a, b \rangle, \quad \mathbb{R}^+, \quad \mathbb{R}, \quad \text{etc.}$$

Beispiel: Zufälliger Punkt im Einheitskreis



Beispiel (Forts.)

sG X ... Abstand des Punktes vom Ursprung

Merkmalraum: $M_X = [0, 1)$

$$P(X \leq x) = \frac{\text{Fläche des Kreises mit Radius } x}{\text{Fläche des Kreises mit Radius } 1} = \frac{x^2 \pi}{\pi} = x^2$$

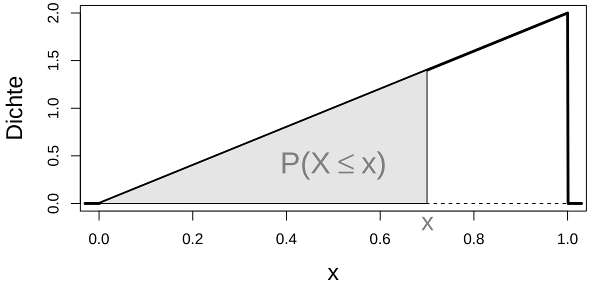
$$\text{für } 0 \leq x < 1$$

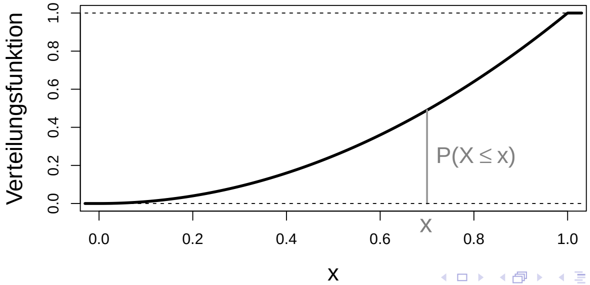
Beispiel (Forts.)

VF: $$F_X(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Dichte: $$f_X(x) = F'_X(x) = \begin{cases} 2x & 0 \leq x < 1 \\ 0 & \text{sonst} \end{cases}$$

Beispiel (Forts.)





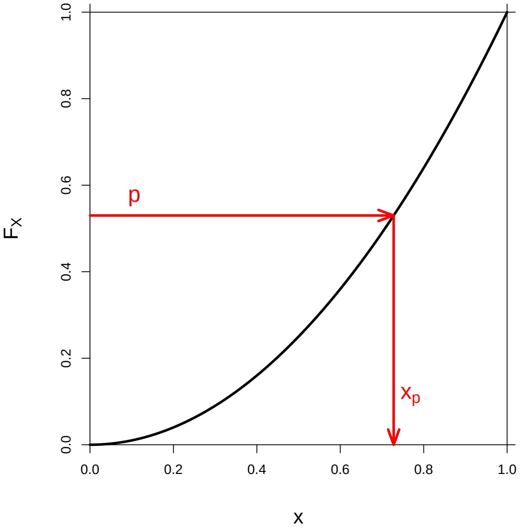
Quantilenbestimmung (Bsp.)

$$F_X(x_p) = P(X \leq x_p) =$$

$$= \int_{-\infty}^{x_p} f_X(t)\,dt = p$$

$$\Longleftrightarrow$$

$$x_p = F_X^{-1}(p)$$



Gemischte sGn

In einigen praktisch wichtigen Situationen ist die sG – oder ihre Verteilung – weder diskret noch stetig.

Beispiele:

- Eine Person kommt zufällig zu einem geregelten Fußgängerübergang. X_1 sei ihre Wartezeit.
- Man betrachte ein Produkt (z. B. eine Glühlampe), das bereits bei der ersten Inbetriebnahme ausfallen kann. X_2 sei die Lebensdauer dieses Produkts.
- Lebensdauertests von langlebigen Produkten werden meist nach einer vorgegebenen Zeitspanne t_0 abgebrochen. Ist X die Lebensdauer des Produkts, wird also tatsächlich $Z = \min\{X, t_0\}$ beobachtet.
- $\cdots$

Gemischte sGn (Forts.)

Diesen Beispielen ist gemeinsam, dass die jeweilige Verteilung prinzipiell stetig ist, aber einzelne Punkte dennoch eine positive Wahrscheinlichkeit haben:

$$P(X_1 = 0) > 0, \quad P(X_2 = 0) > 0, \quad P(Z = t_0) > 0$$

Die Verteilungsfunktion dieser sGn ist keine Treppenfunktion aber auch nicht überall stetig.

Gemischte sGn (Forts.)

Die Verteilungsfunktion F ist eine Mischung aus einer diskreten (F_d) und einer stetigen (F_s) Verteilungsfunktion:

$$F(x) = \alpha F_d(x) + (1-\alpha) F_s(x), \quad 0 \leq \alpha \leq 1$$

Merkmalraum einer gemischten sG:

$$M = \underbrace{\{x_1, x_2, \ldots, x_m\}}_{\text{diskrete Punkte}} \cup \underbrace{\langle a, b \rangle}_{\text{Intervall}}$$

Es kann auch (abzählbar) unendlich viele diskrete Punkte geben.

Gemischte sGn (Forts.)

Die diskreten Punkte x_i haben positive Wahrscheinlichkeiten $p(x_i) > 0$ und es gibt eine Dichte f^* mit Träger $\langle a, b \rangle$, sodass:

$$\sum_{i=1}^{m} p(x_i) + \int\limits_a^b f^*(x)\, dx = 1$$

$$P(X \in B) = \sum_{x_i \in B} p(x_i) + \int\limits_B f^*(x)\, dx$$

Achtung: f^* ist wegen $\int_a^b f^*(x)\, dx < 1$ keine vollständige Dichte!

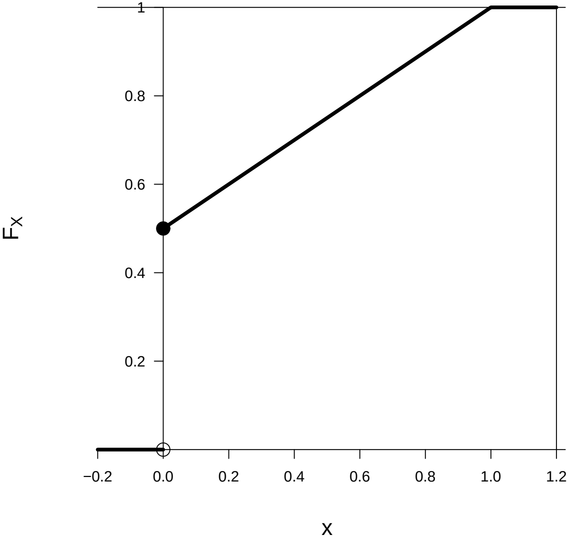
Beispiel

Merkmalraum: $M = \{0\} \cup \langle 0, 1 \rangle$

VF: $$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ \dfrac{x+1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Dichte: $$f^*(x) = F'(x) = \begin{cases} \dfrac{1}{2} & 0 \leq x < 1 \\ 0 & \text{sonst} \end{cases}$$

Beispiel (Forts.)



$$P\left(-3 < X \leq \frac{1}{2}\right) =$$

$$= F\left(\frac{1}{2}\right) - F(-3) =$$

$$= \frac{3}{4} - 0 = \frac{3}{4}$$

$$P(X = 0) =$$

$$= F(0) - F(0-) =$$

$$= \frac{1}{2} - 0 = \frac{1}{2}$$

Mischverteilungen

Allgemeiner spricht man von einer Mischverteilung, wenn sich die Verteilungsfunktion F als Mischung von VFn F_j darstellen lässt:

$$F(x) = \sum_{j=1}^{\infty} \alpha_j F_j(x) \quad \text{mit} \quad \alpha_j > 0 \quad \text{und} \quad \sum_{j=1}^{\infty} \alpha_j = 1$$

Durch Mischen von – endlich oder unendlich vielen – Verteilungen lässt sich die statistische Modellbildung stark erweitern.

Mischverteilungen sind auch in der Bayes'schen Statistik von großer Bedeutung.

Transformation von sGn

Angenommen, man kennt von einer sG X ihre Verteilung:

VF: F_X Dichte: f_X oder   W–Funktion: p_X

Gesucht: Verteilung einer Transformation von X:

$$Y = g(X) \qquad \text{mit} \qquad g : \mathbb{R} \longrightarrow \mathbb{R} \quad \text{(Funktion)}$$

X eine diskrete sG

p_X ... W–Funktion von X

g ... umkehrbar eindeutige Funktion

Gesucht: W–Funktion von $Y = g(X)$

$$p_Y(y) = P(Y = y) = P\big(g(X) = y\big) = P\big(X = g^{-1}(y)\big) = p_X\big(g^{-1}(y)\big)$$

$$\Longrightarrow \quad p_Y(y) = p_X\big(g^{-1}(y)\big)$$

Beispiel 1

X . . . Augenzahl eines Würfels

$$p_X(x) = \frac{1}{6} \quad \text{für} \quad x \in M_X = \{1, 2, 3, 4, 5, 6\}$$

Transformation: $Y = X^2$

$$y = g(x) = x^2 \quad \Longrightarrow \quad g^{-1}(y) = \sqrt{y}$$

$$p_Y(y) = p_X\left(\sqrt{y}\right) = \frac{1}{6} \quad \text{für} \quad y \in M_Y = \{1, 4, 9, 16, 25, 36\}$$

Beispiel 2

Eine Münze wird wiederholt geworfen: T, T, H, T, H, H, ...

X ... Nummer des Wurfes mit dem *ersten* Kopf (H)

$$p_X(x) = \left(\frac{1}{2}\right)^x \quad \text{für} \quad x \in M_X = \{1, 2, \dots\}$$

Transformation:

$$Y = g(X) = \begin{cases} -1 & \text{falls } X \text{ ungerade} \\ 1 & \text{falls } X \text{ gerade} \end{cases}$$

g ist nicht umkehrbar eindeutig! $\longrightarrow$ Direkte Überlegung!

Beispiel 2 (Forts.)

$$p_Y(-1) = P(Y = -1) = P\big(X \in \{1, 3, 5, \dots\}\big) =$$

$$= \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^{2x-1} = \sum_{x=0}^{\infty} \left(\frac{1}{2}\right)^{2x+1} = \frac{1/2}{1 - 1/4} = \frac{2}{3}$$

$$p_Y(1) = 1 - p_Y(-1) = \frac{1}{3}$$

Methode der Verteilungsfunktion

Unabhängig davon, ob die Funktion g umkehrbar eindeutig ist oder nicht, lässt sich für stetiges X die Verteilung von $Y = g(X)$ mittels der Methode der Verteilungsfunktion bestimmen:

- Bestimme die Verteilungsfunktion F_X von X.
- Bestimme die Verteilungsfunktion F_Y von $Y = g(X)$.
- Falls F_Y absolut stetig, bestimme durch Ableiten $f_Y(y) = \dfrac{dF_Y(y)}{dy}$ die Dichte von Y.

Beispiel [29]

X ... Abstand eines zufälligen Punktes im 1–Kreis vom Ursprung

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Transformation: $Y = X^2$; Träger: $\mathcal{S}_Y = (0, 1)$

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P\left(X \leq \sqrt{y}\right) =$$

$$= F_X\left(\sqrt{y}\right) = \left(\sqrt{y}\right)^2 = y$$

Beispiel (Forts.) [29]

$$\Longrightarrow \quad f_Y(y) = F'_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{sonst} \end{cases}$$

Das Quadrat $Y = X^2$ des Abstands X eines zufällig gewählten Punktes im 1–Kreis vom Ursprung ist uniform verteilt.

Transformationssatz

Ist die Funktion g umkehrbar eindeutig, kann die Dichte von $Y = g(X)$ auch direkt bestimmt werden:

X sei eine stetige sG mit Dichte f_X und Träger S_X.

g sei eine umkehrbar eindeutige differenzierbare Funktion auf S_X.

Dann ist die Dichte von $Y = g(X)$ gegeben durch:

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \left|\frac{dg^{-1}(y)}{dy}\right|, \quad y \in S_Y$$

Der Träger S_Y von Y ist die Menge $S_Y = \{y = g(x) \,|\, x \in S_X\}$.

Beweis: Skriptum (S. 129)

Jacobian

Die Ableitung der Umkehrfunktion g^{-1} nennt man häufig auch die Jacobian und schreibt:

$$J = \frac{dx}{dy} = \frac{dg^{-1}(y)}{dy} = \frac{1}{\frac{dg(x)}{dx}}$$

Transformationssatz: $f_Y(y) = f_X\left(g^{-1}(y)\right) |J|, \quad y \in S_Y$

Beispiel (Forts.)

$$y = g(x) = x^2 \quad \Longrightarrow \quad x = g^{-1}(y) = \sqrt{y}$$

$$J = \left(\sqrt{y}\right)' = \frac{1}{2\sqrt{y}}$$

$$\Longrightarrow \quad f_Y(y) = \underbrace{f_X\left(g^{-1}(y)\right)}_{2\sqrt{y}} \underbrace{|J|}_{\frac{1}{2\sqrt{y}}} = 1 \quad \text{für} \quad y \in (0,1)$$

Affine Transformationen

Eine affine Transformation ist eine Abbildung der Form $Y = a + bX$.

Heißt manchmal auch (fälschlich) lineare Transformation.

X sei eine stetige sG mit Dichte f_X.

$Y = a + bX$ sei eine affine Transformation mit $b \neq 0$.

Die Dichte von Y ist gegeben durch:

$$f_Y(y) = \frac{1}{|b|} f_X\left(\frac{y-a}{b}\right)$$

Affine Transformationen (Forts.)

Beweis:

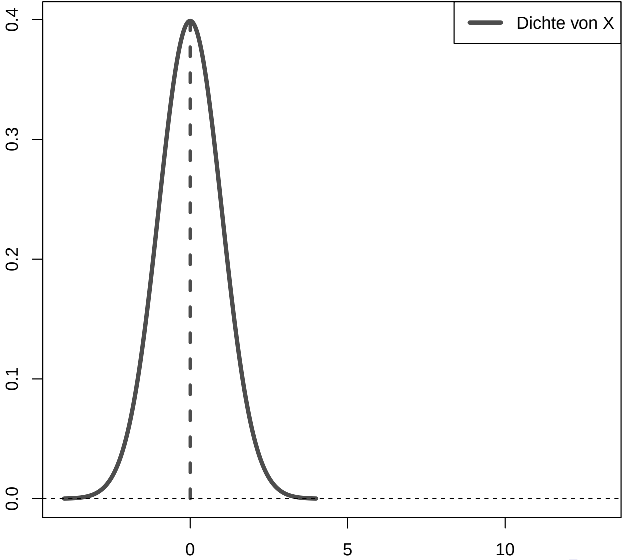
$$y = a + bx \quad \Longrightarrow \quad x = \frac{y-a}{b} \quad \Longrightarrow \quad J = \frac{dx}{dy} = \frac{1}{b}$$

Anwendung des Transformationssatzes:

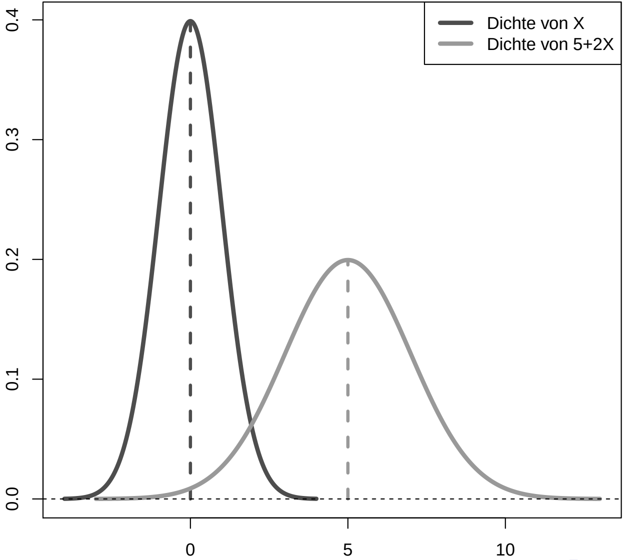
$$f_Y(y) = f_X\left(g^{-1}(y)\right)|J| = \frac{1}{|b|}\, f_X\left(\frac{y-a}{b}\right)$$

Eine affine Transformation ändert die Lage und/oder die Skalierung der Dichte, nicht aber ihre Form.

Beispiel



Beispiel



Erwartungswert

Die W–Funktion (diskreter Fall) oder die Dichte (stetiger Fall) enthält die gesamte Information über eine sG X.

Manchmal genügen einige wenige charakteristische Werte.

Einer dieser Werte ist der Erwartungswert (auch Mittelwert oder kurz Mittel genannt) der sG X, geschrieben $\mathbb{E}(X)$.

Der Erwartungswert ist ein (gewichteter) Durchschnittswert der möglichen Ausprägungen von X.

Erwartungswert einer diskreten sG

X sei eine diskrete sG mit W–Funktion $p(x)$.

Es gelte: $\sum_x |x|p(x) < \infty$ (absolut konvergent)

Der Erwartungswert von X ist definiert durch:

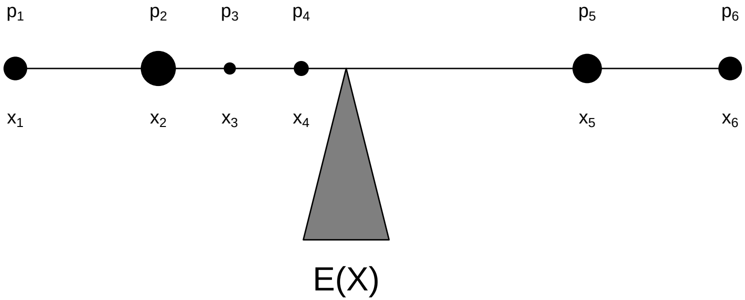
$$\mathbb{E}(X) = \sum_x x\, p(x)$$

Der Erwartungswert von X ist ein gewichteter Mittelwert der möglichen Ausprägungen von X.

Die Gewichte entsprechen den Wahrscheinlichkeiten der Ausprägungen.

Erwartungswert als Schwerpunkt

Der Erwartungswert lässt sich auch als Schwerpunkt von Punktmassen interpretieren.



Beispiel

X ... Augenzahl eines Würfels

Merkmalraum: $M = \{1, 2, 3, 4, 5, 6\}$

W–Funktion: $p(x) = \frac{1}{6}, \; x \in M$

Erwartungswert:

$$\mathbb{E}(X) = \sum_{x=1}^{6} x \left(\frac{1}{6}\right) = \frac{1}{6} \sum_{x=1}^{6} x = \frac{1}{6} \cdot \frac{(1+6)(6)}{2} = \frac{7}{2} = 3.5$$

Bem: Der Erwartungswert einer sG X ist nicht notwendigerweise ein Element des Merkmalraums M_X !

Beispiel (Forts.)

Der Erwartungswert lässt sich auch wie folgt interpretieren:

Angenommen, man wirft den Würfel n Mal, wobei n eine große Zahl sei, z. B. $n = 1000$.

Ist $H_n(x)$ die (absolute) Häufigkeit der Augenzahl x, so beträgt die durchschnittliche geworfene Augenzahl:

$$\frac{1}{n}\sum_{x=1}^{6} x\, H_n(x) = \sum_{x=1}^{6} x \underbrace{\frac{H_n(x)}{n}}_{\approx\, p(x)} \approx \sum_{x=1}^{6} x\, p(x) = \mathbb{E}(X)$$

Dabei wird die frequentistische Interpretation von Wahrscheinlichkeit zugrunde gelegt.

Erwartungswert einer stetigen sG

X sei eine stetige sG mit Dichte $f(x)$.

Es gelte: $\int\limits_{-\infty}^{\infty} |x| f(x)\, dx < \infty$ (absolut konvergent)

Der Erwartungswert von X ist definiert durch:

$$\mathbb{E}(X) = \int\limits_{-\infty}^{\infty} x\, f(x)\, dx$$

Beispiel

Die Dichte der sG X sei gegeben durch:

$$f(x) = \begin{cases} 4x^3 & 0 < x < 1 \\ 0 & \text{sonst} \end{cases}$$

Erwartungswert:

$$\mathbb{E}(X) = \int_0^1 x\,(4x^3)\,dx = \int_0^1 4x^4\,dx = \frac{4x^5}{5}\Big|_0^1 = \frac{4}{5}$$

Erwartungswert einer gemischten Verteilung

Merkmalraum: $M = \{x_1, x_2, \ldots, x_m\} \cup \langle a, b \rangle$

Punktwahrscheinlichkeiten / Dichte: $p(x_i)$, $f^*(x)$

Erwartungswert:

$$\mathbb{E}(X) = \sum_{i=1}^{m} x_i \, p(x_i) + \int_a^b x \, f^*(x) \, dx$$

Dabei wird vorausgesetzt, dass Summe (falls $m = \infty$) und Integral absolut konvergieren.

Beispiel [38]

$$p(0) = \frac{1}{2}, \quad f^*(x) = \frac{1}{2},\ 0 < x < 1$$

Erwartungswert:

$$\mathbb{E}(X) = (0)\left(\frac{1}{2}\right) + \int_0^1 \frac{x}{2}\,dx = \left.\frac{x^2}{4}\right|_0^1 = \frac{1}{4}$$

Erwartungswert einer Funktion von X (diskreter Fall)

X sei eine diskrete sG mit W–Funktion p_X.

Die sG $Y = g(X)$ sei eine Funktion von X.

Es gelte: $\sum_{x \in S_X} |g(x)| p_X(x) < \infty$

Dann existiert der Erwartungswert von Y und ist gegeben durch:

$$\mathbb{E}(Y) = \sum_{x \in S_X} g(x) p_X(x)$$

Erwartungswert einer Funktion von X (stetiger Fall)

X sei eine stetige sG mit Dichte f_X.

Die sG $Y = g(X)$ sei eine Funktion von X.

Es gelte: $\int\limits_{-\infty}^{\infty} |g(x)| f_X(x)\, dx < \infty$

Dann existiert der Erwartungswert von Y und ist gegeben durch:

$$\mathbb{E}(Y) = \int\limits_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

Erwartungswert einer Funktion von X (Forts.)

Ein analoger Satz gilt auch für den gemischten Fall:

$$\mathbb{E}\big[g(X)\big] = \sum_{i=1}^{m} g(x_i) p_X(x_i) + \int_{-\infty}^{\infty} g(x) f_X^*(x)\, dx$$

Diese Aussagen werden auch als Law of the Unconscious Statistician (kurz: LotUS) bezeichnet.

Der LotUS wird häufig als Definition für den Erwartungswert einer Funktion $Y = g(X)$ verwendet, ist aber ein (mathematischer) Satz.

Beispiel [38]

Der Erwartungswert von z. B. $Y = \left(X - \frac{1}{4}\right)^2$ lässt sich wie folgt berechnen:

$$\mathbb{E}(Y) = \underbrace{\left(0 - \frac{1}{4}\right)^2 \left(\frac{1}{2}\right)}_{\frac{1}{32}} + \underbrace{\int_0^1 \left(x - \frac{1}{4}\right)^2 \left(\frac{1}{2}\right) dx}_{\frac{7}{96}} = \frac{5}{48}$$

Der Erwartungswert von Y ist die Varianz der sG X (wird später noch ausführlicher behandelt).

Vorteil des LotUS

Der Vorteil des LotUS besteht darin, dass man nicht zuerst die Verteilung (Dichte, W–Funktion) von $Y = g(X)$ bestimmen muss, wenn man den Erwartungswert von Y berechnen möchte, sondern mit der Verteilung (Dichte, W–Funktion) von X arbeiten kann.

Stetiger Fall:

$$\text{nach Definition:} \quad \mathbb{E}(Y) = \int\limits_{-\infty}^{\infty} y\, f_Y(y)\, dy$$

$$\text{mit LotUS:} \quad \mathbb{E}(Y) = \int\limits_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

Die zweite Berechnungsmöglichkeit ist häufig einfacher.

Eigenschaften des Erwartungswerts

Für Konstanten a, b, k_1, k_2 und Funktionen g, h gilt:

(1) $\mathbb{E}(a) = a$

(2) $\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b$

(3) $\mathbb{E}\big[k_1 g(X) + k_2 h(X)\big] = k_1 \mathbb{E}\big[g(X)\big] + k_2 \mathbb{E}\big[h(X)\big]$

Median

Neben dem Erwartungswert ist der Median (= 50% Quantil) einer sG X ein zweiter wichtiger Lageparameter:

$$\tilde{x} = \text{med}(X) = F_X^{-1}\left(\frac{1}{2}\right)$$

$$F^{-1} \quad \ldots \text{ verallgemeinerte Inverse von } F_X$$

Bei um einen Punkt a symmetrischen Verteilungen, stimmen beide Lageparameter überein:

$$\mathbb{E}(X) = \text{med}(X)$$

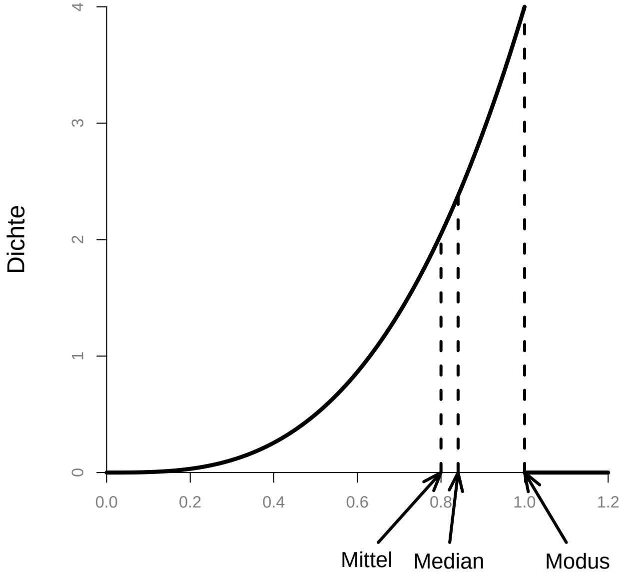
Beispiel [62]

$$f(x) = \begin{cases} 4x^3 & 0 < x < 1 \\ 0 & \text{sonst} \end{cases} \quad \Longrightarrow \quad F(x) = \begin{cases} 0 & x \le 0 \\ x^4 & 0 < x < 1 \\ 1 & x \ge 1 \end{cases}$$

Erwartungswert: $\mathbb{E}(X) = \int_0^1 x\left(4x^3\right) dx = \frac{4}{5} = 0.8$

Median: $\widetilde{x} = F^{-1}\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{1/4} \doteq 0.84$

Beispiel (Forts.)



Varianz

Neben Maßzahlen für die Lage benötigt man auch Maßzahlen für das Streuungsverhalten einer Verteilung. Die wichtigste Maßzahl dieser Art ist die Varianz.

Ist X eine sG mit Mittelwert $\mu_X = \mathbb{E}(X)$, sodass $\mathbb{E}\left[(X-\mu_X)^2\right] < \infty$, dann ist die Varianz von X definiert durch:

$$\sigma_X^2 = \mathrm{Var}(X) = \mathbb{E}\left[(X-\mu_X)^2\right]$$

Die positive Wurzel aus der Varianz ist die Streuung:

$$\sigma_X = +\sqrt{\sigma_X^2} = +\sqrt{\mathrm{Var}(X)}$$

Varianz (Forts.)

Die Varianz ist die mittlere quadratische Abweichung einer sG von ihrem Mittelwert, somit der Erwartungswert von $Y = g(X) = (X - \mu)^2$. Nach dem LotUS gilt:

$$\text{diskret:} \quad \mathrm{Var}(X) = \sum_x \left[x - \mathbb{E}(X)\right]^2 p_X(x)$$

$$\text{stetig:} \quad \mathrm{Var}(X) = \int_{-\infty}^{\infty} \left[x - \mathbb{E}(X)\right]^2 f_X(x)\, dx$$

$$\text{gemischt:} \quad \mathrm{Var}(X) = \sum_{i=1}^{m} \left[x_i - \mathbb{E}(X)\right]^2 p(x_i) + \int_{-\infty}^{\infty} \left[x - \mathbb{E}(X)\right]^2 f^*(x)\, dx$$

Verschiebungssatz

Die Varianz σ_X^2 einer sG X lässt sich auch wie folgt berechnen:

$$\sigma_X^2 = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \mathbb{E}(X^2) - \mu_X^2$$

Beweis:

$$\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}(X^2 - 2\mu X + \mu^2) =$$
$$= \mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(X^2) - \mu^2$$

Folgerung: $\sigma_X^2 \geq 0 \quad \Longrightarrow \quad \mathbb{E}(X^2) \geq \mathbb{E}^2(X)$

Beispiel 1 [62]

$$\mathbb{E}(X) = \frac{4}{5}, \qquad \mathrm{Var}(X) = \int_0^1 \left(x - \frac{4}{5}\right)^2 (4x^3)\,dx$$

Einfacher ist die Berechnung mittels Verschiebungssatz:

$$\mathbb{E}(X^2) = \int_0^1 x^2(4x^3)\,dx = 4\int_0^1 x^5\,dx = \left.\frac{2x^6}{3}\right|_0^1 = \frac{2}{3}$$

$$\Longrightarrow \quad \mathrm{Var}(X) = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75} \quad \Longrightarrow \quad \sigma_X = \sqrt{\frac{2}{75}}$$

Beispiel [38]

$$p(0) = \frac{1}{2}, \quad f^*(x) = \frac{1}{2}, \; 0 < x < 1$$

Verschiebungssatz:

$$\mathbb{E}(X^2) = (0^2)\left(\frac{1}{2}\right) + \int_0^1 \frac{x^2}{2}\, dx = \frac{x^3}{6}\bigg|_0^1 = \frac{1}{6}$$

$$\Longrightarrow \quad \mathrm{Var}(X) = \frac{1}{6} - \left(\frac{1}{4}\right)^2 = \frac{10}{96} = \frac{5}{48}$$

Eigenschaften der Varianz/Streuung

Für Konstanten a, b gilt:

(1) $\mathrm{Var}(a) = 0$

(2) $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$

(3) $\sigma_{aX+b} = a\,\sigma_X$

Einheiten der Kenngrößen

Bei konkreten Anwendungen ist zu beachten, dass den hier behandelten Kenngrößen μ_X (Mittelwert), $\mathrm{med}(X)$ (Median), σ_X^2 (Varianz) und σ_X (Streuung) physikalische Einheiten zugeordnet sind.

Bsp: Ist beispielsweise die sG X ein Gewicht in der Einheit [kg], haben die Größen μ_X, $\mathrm{med}(X)$ (und andere Quantile) und σ_X die Einheit [kg].

σ_X^2 hat hier die – schwierig zu interpretierende – Einheit [kg^2].

Aus diesem Grund gibt man in Anwendungen der Streuung meist den Vorzug gegenüber der Varianz.

MAD

Ein alternatives Streumaß ist die mittlere absolute Abweichung (kurz MAD) der sG X vom Median $\text{med}(X)$.

Für eine stetige sG X mit Dichte f_X ist der MAD definiert durch:

$$\text{MAD}(X) = \int_{-\infty}^{\infty} \left| x - \text{med}(X) \right| f_X(x)\, dx$$

Bsp [62]:

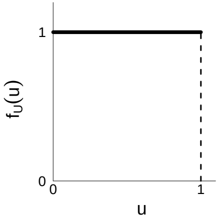
$$\text{MAD}(X) = \int_0^1 \left| x - \left(\frac{1}{2}\right)^{1/4} \right| 4x^3\, dx = \frac{5}{16}$$

Uniform verteilte Zufallszahlen

Die Simulation von stochastischen Vorgängen verschiedenster Art ist mittlerweile ein unverzichtbares Werkzeug der modernen (Computer–) Statistik.

Dazu benötigt man unbegrenzt viele Realisationen einer sG U mit der folgenden uniformen Dichte:

$$f_U(u) = \begin{cases} 1 & 0 < u < 1 \\ 0 & \text{sonst} \end{cases}$$



Pseudozufallszahlen: In der Praxis verwendet man meist algorithmisch erzeugte Zahlen, die den Anschein von echten Zufallszahlen erwecken.

R: runif()

Erzeugung von 10000 uniform verteilten (Pseudo–) Zufallszahlen.

Darstellung des Ergebnisses in Form eines (flächentreuen) Histogramms.

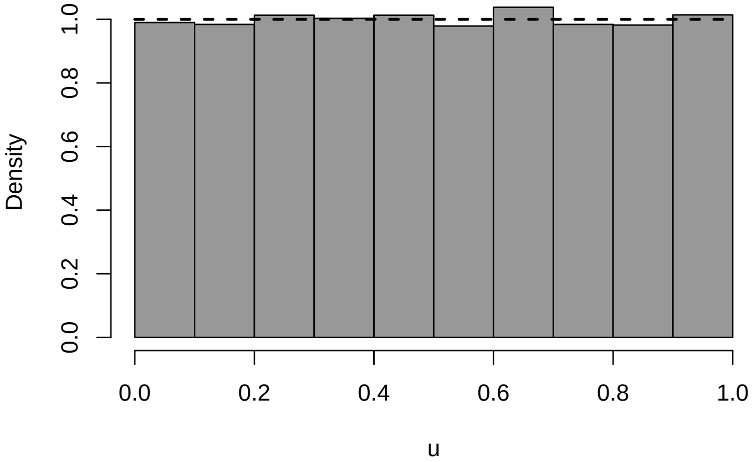
```
Nsim <- 10^4

u <- runif(Nsim)

hist(u, breaks=(0:10)/10, prob=TRUE, main="", col="grey60")

lines(c(0,1), c(1,1), lty=2, lwd=2)
```

Histogramm von 10^4 mit `runif()` erzeugten Zufallszahlen



Bezeichnung

Hat die sG X die Verteilungsfunktion F_X, oder die Dichte f_X, oder die W–Funktion p_X, so schreibt man häufig:

$$X \sim F_X, \qquad X \sim f_X, \qquad X \sim p_X$$

sprich: "X ist verteilt nach ..."

Ist keine Verwechslung möglich, schreibt man:

$$X \sim F, \qquad X \sim f, \qquad X \sim p$$

Integraltransformation

X sei eine stetige sG mit Verteilungsfunktion F.

Dann ist die sG $Y = F(X)$ uniform verteilt:

$$Y = F(X) \sim \text{uniform}$$

Der Name der Transformation kommt daher, dass im stetigen Fall die Verteilungsfunktion F das Integral über die Dichte f ist:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

Inverse Integraltransformation

F sei eine beliebige Verteilungsfunktion (diskret, stetig oder gemischt).

U sei eine sG mit uniformer Dichte $f_U(u) = I_{(0,1)}(u)$.

Dann ist die sG $X := F^{-1}(U)$ nach F verteilt:

$$X = F^{-1}(U) \sim F$$

$$F^{-1} \ \ldots \ \text{verallgemeinerte Inverse von } F$$

Beweisidee: Skriptum (S. 141)

Inversionsmethode

Man möchte eine Realisation x einer sG $X \sim F$ generieren.

Dazu genügen die beiden folgenden Schritte:

(1) Erzeuge eine Realisation u von $U \sim f_U(u) = I_{(0,1)}(u)$.

(2) Bilde $x = F^{-1}(u)$.

Beispiel (stetiger Fall)

Man möchte Realisationen einer sG X mit Dichte f generieren:

$$X \sim f(x) = e^{-x} I_{(0,\infty)}(x)$$

Zuerst bestimmt man die Verteilungsfunktion von X:

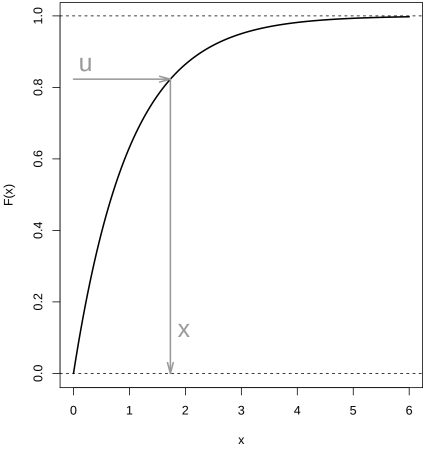
$$F(x) = \int_0^x e^{-t}\, dt = 1 - e^{-x}, \quad 0 < x < \infty$$

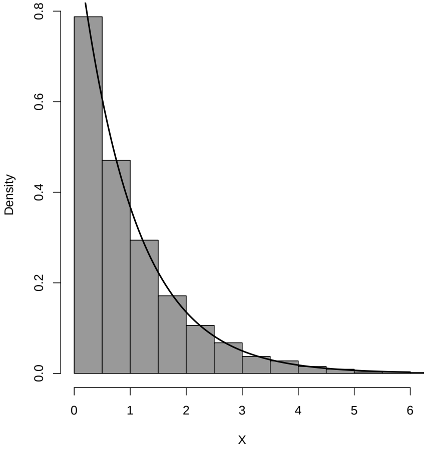
F ist zu invertieren:

$$1 - e^{-x} = u \quad \Longrightarrow \quad x = -\ln(1-u)$$

$$u \ \ldots \ \text{uniform verteilte Zufallszahl}$$

Beispiel (stetiger Fall) (Forts.)





Beispiel (diskreter Fall)