

Stetige Zufallsvariable

Wir nennen eine Zufallsvariable X stetig, wenn ihre kumulative Verteilungsfunktion $F(x) = P(X \leq x)$ stetig ist.

Dichtefunktion (pdf) einer stetigen Zufallsvariable X

Die stetige Zufallsvariable X besitzt eine (stückweise) differenzierbare Verteilungsfunktion $F(x)$. Wir nennen

$$f(x) = F'(x)$$

die Dichtefunktion. Zudem gilt:

$$F(x) = \int_{-\infty}^x f(s) ds.$$

Zusammenhang pdf und cdf einer stetigen Zufallsvariable

Siehe <https://demonstrations.wolfram.com/ConnectingTheCDFAndThePDF/>.

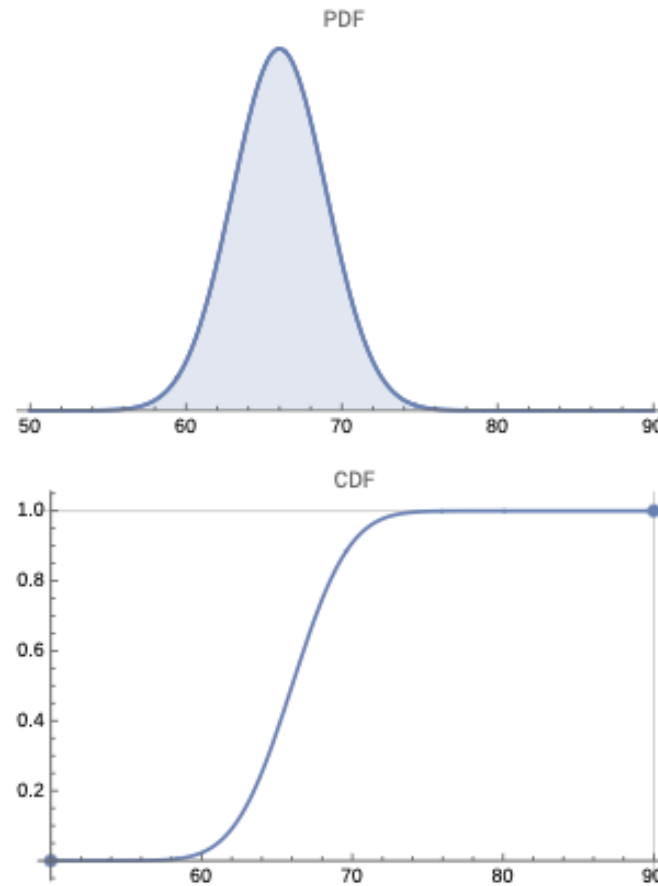


Abb.: Oben die pdf und unten die cdf.

Anmerkungen zur pdf einer stetigen Zufallsvariable

Sei X eine stetige Zufallsvariable und $f(x)$ ihre Dichtefunktion.

- ▶ Die Dichtefunktion ist normiert und es gilt $\int_{-\infty}^{+\infty} f(s)ds = 1$.
- ▶ $f(x)$ ist auf ganz \mathbb{R} definiert und nimmt nur nichtnegative Werte an.
- ▶ Wir nennen $\{x \in \mathbb{R} : f(x) > 0\}$ den Träger der Dichtefunktion f .
- ▶ Es gilt für stetige Zufallsvariablen: $P(X = x_i) = 0$. Die Punktwahrscheinlichkeit ist demnach stets 0.

Anmerkungen zur cdf einer stetigen Zufallsvariable

Sei X eine stetige Zufallsvariable und $F(x)$ ihre Verteilungsfunktion.

- ▶ Die cdf ist monoton steigend und es gilt:
 - ▶ $\lim_{x \rightarrow -\infty} F(x) = 0$.
 - ▶ $\lim_{x \rightarrow \infty} F(x) = 1$.
 - ▶ Für $x < y$: $F(x) \leq F(y)$.
- ▶ Falls die cdf an einer Stelle springt, entspricht die Höhe des Sprungs genau der Wahrscheinlichkeit, mit welcher das Ereignis an jener Stelle eintritt.
- ▶ Die Verteilungsfunktion $F(x)$ ist rechtsseitig stetig.

Berechnung von Quantilen

Sei X eine stetige oder diskrete Zufallsvariable. Wir nennen den Wert $x_p \in \mathbb{R}$ mit $F(x_p) = p$ das p -Quantil von X .

Intervallswahrscheinlichkeit einer stetigen Zufallsvariable

Sei X eine stetige Zufallsvariable und die Wahrscheinlichkeit eines Ereignisses im Intervall $[s_1, s_2]$ ist

$$P(s_1 \leq x \leq s_2) = F(s_2) - F(s_1) = \int_{s_1}^{s_2} f(s) ds.$$

Anmerkung: Analoges gilt für $[s_1, s_2)$, $(s_1, s_2]$ oder (s_1, s_2) . Die rechte Seite von oben bleibt dieselbe, da die Wahrscheinlichkeit $P(X = s_i)$ für $i \in \{1, 2\}$ einer stetigen Zufallsvariable Null ist.

Erwartungswert von Zufallsvariablen

Sei X eine diskrete oder stetige Zufallsvariable. Der Erwartungswert von X wird wie folgt berechnet:

$$\mu = E(X) = \begin{cases} \sum_{i=1}^n x_i \cdot p_i & , \text{ für } X \text{ diskret,} \\ \int_{-\infty}^{+\infty} x \cdot f(x) dx & , \text{ für } X \text{ stetig.} \end{cases}$$

Zudem gilt für $a, b \in \mathbb{R}$ und X und Y Zufallsvariablen:

$$E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y).$$

Falls die Zufallsvariablen X und Y unabhängig sind, gilt

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

Für eine Zufallsvariable X , die eine symmetrische Dichtefunktion um einen Wert m hat, gilt: $E(X) = m$.

Erwartungswert einer Funktion $g(\cdot)$

Sei X eine diskrete oder stetige Zufallsvariable und $g(\cdot)$ eine beliebige reelle Funktion. Es gilt:

$$E(g(X)) = \begin{cases} \sum_{i=1}^n g(x_i) \cdot p_i & , \text{ für } X \text{ diskret,} \\ \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx & , \text{ für } X \text{ stetig.} \end{cases}$$

Beispiel 1: Ein Zufallsgenerator erzeugt eine reelle Zahl zwischen 0 und 4. Die Dichtefunktion der Zufallsvariable ist:

$$f(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{1}{4} & \text{für } 0 \leq x \leq 4 \\ 0 & \text{für } x > 4. \end{cases}$$

Berechne den Erwartungswert.

Lösung Beispiel 1:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{-\infty}^0 x \cdot 0 dx + \int_0^4 x \cdot \frac{1}{4} dx + \int_4^{+\infty} x \cdot 0 dx \\ &= \left[\frac{1}{4} \cdot \frac{x^2}{2} \right]_0^4 \\ &= \frac{16}{8} - 0 \\ &= 2. \end{aligned}$$

Beispiel 2: Pro Jahr treten folgende Anzahl an Hitzetage mit entsprechenden Wahrscheinlichkeiten auf.

Anzahl	0	1	2
p_i	0.1	0.5	0.4

Dies verursacht folgende Kosten in tausend Euro bei einem Bauunternehmer:

$$K(x) = 300 - \frac{100}{5 + 2 \cdot x}.$$

Was sind die zu erwartenden Kosten pro Jahr?

Lösung Beispiel 2:

$$\begin{aligned} E(K(X)) &= \sum_{i=0}^2 K(x_i) \cdot p_i \\ &= 0.1 \cdot \left(300 - \frac{100}{5 + 2 \cdot 0}\right) + 0.5 \cdot \left(300 - \frac{100}{5 + 2 \cdot 1}\right) + 0.4 \cdot \left(300 - \frac{100}{5 + 2 \cdot 2}\right) \\ &= 286.41 \text{ tsd Euro} \end{aligned}$$

Varianz von Zufallsvariablen

Sei X eine diskrete oder stetige Zufallsvariable mit $E(X) = \mu$. Die Varianz von X berechnet sich wie folgt:

$$\sigma^2 = \text{Var}(X) = \begin{cases} \sum_{i=1}^n (x_i - \mu)^2 \cdot p_i & , \text{ für } X \text{ diskret,} \\ \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) dx & , \text{ für } X \text{ stetig.} \end{cases}$$

Zudem gilt für $a, b \in \mathbb{R}$ das Folgende:

$$\text{Var}(a \cdot X + b) = a^2 \cdot \text{Var}(X).$$

Außerdem gilt:

- ▶ $\text{Var}(X) = E((X - \mu)^2)$
- ▶ $E((X - \mu)^2) = E(X^2) - \mu^2.$

Kovarianz von zwei Zufallsvariablen

Seien X und Y Zufallsvariablen mit Erwartungswerten $E(X) = \mu_X$ und $E(Y) = \mu_Y$. Die Kovarianz ist wie folgt definiert:

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)).$$

Für $a, b \in \mathbb{R}$ gilt:

$$\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y).$$

Zwei Zufallsvariablen X und Y heißen unkorreliert, wenn $\text{Cov}(X, Y) = 0$ ist.

Falls X und Y unabhängig sind, gilt ebenfalls

$$\text{Cov}(X, Y) = 0.$$

Demnach sind unabhängige Zufallsvariablen auch unkorreliert. Aus der Unkorreliertheit folgt aber nicht zwingend, dass die Zufallsvariablen unabhängig sind.

Anmerkung: Die Kovarianz ist ein Maß für die gegenseitige Abhängigkeit von X und Y . Sie gibt die Richtung aber nicht die Stärke des Zusammenhangs an.

Korrelation von zwei Zufallsvariablen

Um die Stärke des Zusammenhangs zu bestimmen, wird die Korrelation verwendet.

Für Zufallsvariablen X und Y mit Erwartungswert $E(X) = \mu_X$ und $E(Y) = \mu_Y$ sowie Varianz $Var(X) = \sigma_X^2$ und $Var(Y) = \sigma_Y^2$ wird die Korrelation wie folgt berechnet:

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

Anmerkung: Die Standardabweichung σ ergibt sich als Wurzel aus der Varianz σ^2 .

Schiefe

Die Schiefe gibt an, ob und wie stark die Verteilung sich nach rechts oder nach links neigt.

Sei X eine Zufallsvariable mit Erwartungswert μ und Standardabweichung σ . Die Schiefe (auch drittes zentrales Moment genannt) ist wie folgt definiert:

$$\gamma = E\left(\left(\frac{X - \mu}{\sigma}\right)^3\right).$$

Für eine symmetrische Verteilung gilt: $\gamma = 0$. Falls $\gamma < 0$ liegt eine linksschiefe Verteilung vor. Bei $\gamma > 0$ liegt eine rechtsschiefe Verteilung vor.

Es gilt:

- ▶ **Rechtsschief** ist identisch mit dem Begriff **linkssteil**.
- ▶ **Linksschief** ist identisch mit dem Begriff **rechtssteil**.

Wölbung

Die Wölbung (auch Kurtosis genannt) gibt die Steilheit einer Verteilung an. Sie wird für eine Zufallsvariable X mit Erwartungswert μ und Standardabweichung σ wie folgt berechnet:

$$\kappa = E\left(\left(\frac{X - \mu}{\sigma}\right)^4\right).$$

Anmerkung: Für eine Normalverteilung gilt $\kappa = 3$. Falls $\kappa > 3$ ist die vorliegende Verteilung spitzer als die Normalverteilung. Für $\kappa < 3$ ist sie flachgipfliger.

Wichtige Wahrscheinlichkeitsverteilungen

Die wichtigsten Verteilungen lernen wir im **DataCamp Kurs** kennen. Diese sind unter anderem:

- ▶ Diskrete Gleichverteilung
- ▶ Geometrische Verteilung
- ▶ Normalverteilung
- ▶ Poisson-Verteilung
- ▶ t-Verteilung
- ▶ Exponentialverteilung
- ▶ ...

Identisch verteilte Zufallsvariablen

Seien X und Y Zufallsvariablen. Sie heißen **identisch verteilt**, wenn beide dieselbe Verteilungsfunktion (cdf) haben. Demnach muss $\forall x \in \mathbb{R}$ gelten:

$$P(X \leq x) = P(Y \leq x).$$

Für identisch verteilte Zufallsvariablen X_1, \dots, X_n mit

$$\mu = E(X_1) = \dots = E(X_n)$$

ist der Erwartungswert des arithmetischen Mittels $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ gegeben durch

$$\begin{aligned} E(\bar{X}) &= \frac{1}{n}(E(X_1) + \dots + E(X_n)) \\ &= \frac{1}{n}(\mu + \dots + \mu) \\ &= \mu. \end{aligned}$$

i.i.d.

X_1, \dots, X_n werden **i.i.d.** (engl: independent and identically distributed) genannt, falls sie unabhängig und identisch verteilt sind.

Das Gesetz der großen Zahlen

Seien X_1, \dots, X_n i.i.d. Zufallsvariablen mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$. Das arithmetische Mittel der Zufallsvariablen berechnet sich wie folgt: $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ und ist ebenfalls eine Zufallsvariable. Für ein beliebiges $\epsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| < \epsilon) = 1.$$

Beispiel:

Anzahl Würfe faire Münze	10	100	1000	10000
Relative Häufigkeit Zahl	0.35	0.42	0.52	0.49

Achtung: Es gibt kein Gesetz des Ausgleichs. Falls zum Beispiel beim Roulette häufig die Farbe schwarz der Ausgang war, sagt das Gesetz der großen Zahlen nicht, dass die Farbe rot ihren Rückstand aufholt. Es handelt sich in jeder Runde um unabhängige Experimente.

Der zentrale Grenzwertsatz

Sei X_1, X_2, \dots eine Folge von i.i.d. Zufallsvariablen mit $E(X_i) = \mu$ und $\text{Var}(X_i) = \sigma^2$. Die n -te Teilsumme dieser Folge ist definiert als $S_n = X_1 + X_2 + \dots + X_n$ und es gilt:

$$\begin{aligned} E(S_n) &= n \cdot \mu, \\ \text{Var}(S_n) &= n \cdot \sigma^2. \end{aligned}$$

Dann ist die standardisierte n -te Teilsumme definiert als

$$Z_n = \frac{S_n - n \cdot \mu}{\sqrt{n \cdot \sigma^2}},$$

mit $E(Z_n) = 0$ und $\text{Var}(Z_n) = 1$. Ist Φ die Verteilungsfunktion der Standardnormalverteilung $N(0, 1)$, dann gilt für beliebiges $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P(Z_n \leq x) = \Phi(x).$$

Anmerkung: Unerheblich welche Verteilung X_1, X_2, \dots hat, die n -te Teilsumme ist für hinreichend großes n standardnormalverteilt.