# Statistik I für Betriebswirte
# Vorlesung 3

Dr. Andreas Wünsche

TU Bergakademie Freiberg
Institut für Stochastik

15. April 2019

# Verteilungsfunktion einer Zufallsgröße

- Die Verteilungen von diskreten oder stetigen Zufallsgrößen (oder anderen Typen) können vollständig durch die Verteilungsfunktion der jeweiligen Zufallsgröße beschrieben werden.

- **Definition:** Die Funktion $F_X$ einer reellen Variablen mit reellen Funktionswerten, die durch

$$F_X(x) = P(X < x) = P(-\infty < X < x), \qquad x \in \mathbb{R},$$

  definiert wird, heißt Verteilungsfunktion der Zufallsgröße $X$.

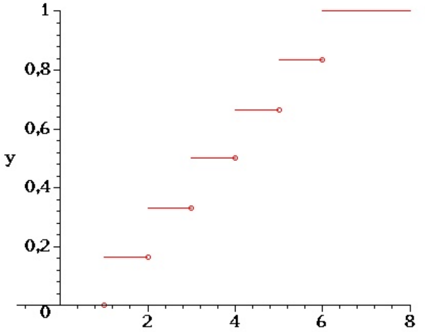
  Der Funktionswert ist für jede reelle Zahl $x$ die Wahrscheinlichkeit dafür, dass die Zufallsgröße $X$ einen Wert annimmt, der kleiner als $x$ ist.

- **Bemerkung:** Mitunter wird die Verteilungsfunktion einer Zufallsgröße $X$ auch durch $\widetilde{F}_X(x) = P(X \le x)$, $x \in \mathbb{R}$, definiert, insbesondere in der Zuverlässigkeitstheorie.

# Verteilungsfunktion einer diskreten Zufallsgröße

- Für diskrete Zufallsgrößen $X$ mit endlich oder auch unendlich vielen möglichen Werten $x_i$ ist die Verteilungsfunktion eine Treppenfunktion mit Sprüngen der Höhe $p_i = P(X = x_i)$ an den Werten $x_i$.

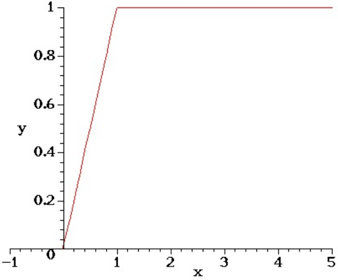
- **Beispiel 1.1:** $X$ Augenzahl beim Würfeln mit einem gerechten Würfel: Verteilungsfunktion $F_X$.



# Verteilungsfunktion einer stetigen Zufallsgröße

- Für stetige Zufallsgrößen ist die Verteilungsfunktion eine in allen Punkten stetige Funktion.
- $$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t, \quad x \in \mathbb{R} \qquad \text{und} \quad f_X(x) = F_X'(x)$$

  in den Werten $x$, in denen die Ableitung existiert.
- **Beispiel 1.2:** Zufallsgröße $X$ auf $[0,1]$ gleichverteilt: Verteilungsfunktion $F_X$.



# Allgemeine Eigenschaften von Verteilungsfunktionen

- Eine Verteilungsfunktion $F_X$ ist monoton nicht fallend.
- Es gilt $\lim\limits_{x \to -\infty} F_X(x) = 0\,.$
- Es gilt $\lim\limits_{x \to +\infty} F_X(x) = 1\,.$
- Es gilt für beliebige reelle Zahlen $a < b$:

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

- Spezialfälle:

$$\begin{aligned} a = -\infty: \qquad P(X < b) &= F_X(b)\,, \\ b = \infty: \qquad P(a \leq X) &= 1 - F_X(a)\,. \end{aligned}$$

- Für eine stetige Zufallsgröße $X$ gelten

$$P(a \leq X < b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X \leq b)\,.$$

# Quantile einer stetigen Zufallsgröße

- Die reelle Zahl $x_q$ mit $0 < q < 1$ heißt $q-$Quantil der stetigen Zufallsgröße $X$, wenn die Werte von $X$ mit einer Wahrscheinlichkeit $q$ links von $x_q$ liegen, d.h. $x_q$ ist eine Lösung der Gleichung

$$P(X < x_q) = F_X(x_q) = q \quad \Longrightarrow \quad x_q = F_x^{-1}(q)\,.$$

- $q-$Quantile können auch für diskrete und andere Zufallsgrößen betrachtet werden.

- Wichtige Quantile sind:
  - das 0.5–Quantil, es heißt Median von $X$;
  - das 0.25– bzw. 0.75–Quantil, dies sind die sogenannten Viertelquantile (Quartile) von $X$ ;
  - die $\alpha-$, $(1-\alpha)-$, $\left(1-\dfrac{\alpha}{2}\right)-$ Quantile für kleine Werte $\alpha$ ,

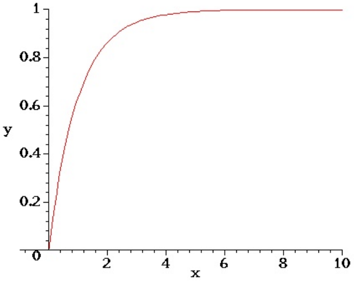
    sie spielen bei statistischen Fragen eine große Rolle.

# Exponentialverteilung

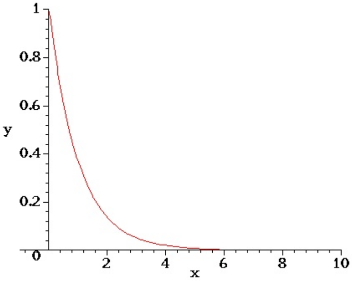
Eine Zufallsgröße $X$ heißt exponentialverteilt mit Parameter $\lambda > 0$, falls für die Verteilungsfunktion $F_X$ bzw. die Verteilungsdichte $f_X$ gilt:

$$F_X(x) = \begin{cases} 0\,, & x \leq 0\,, \\ 1 - \exp(-\lambda x)\,, & x > 0\,, \end{cases} \qquad f_X(x) = \begin{cases} 0\,, & x \leq 0\,, \\ \lambda \exp(-\lambda x)\,, & x > 0\,. \end{cases}$$

Verteilungsfunktion ($\lambda = 1$)




Dichtefunktion ($\lambda = 1$)



# Quantile für Exponentialverteilung

- Es sei $X$ exponentialverteilt mit Parameter $\lambda = 1$, d.h.

$$F_X(x) = P(X < x) = \begin{cases} 0, & x \leq 0, \\ 1 - \exp(-x), & x > 0. \end{cases}$$

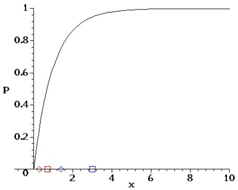
- Dann gilt für das $q-$Quantil $x_q$ (mit $0 < q < 1$) :

$$F_X(x_q) = 1 - \exp(-x_q) = q, \quad \text{also} \quad x_q = -\ln(1-q)\,.$$

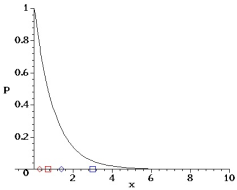
- 

| $q$ | $x_q$ |
|---|---|
| 0.25 | 0.288 |
| 0.5 | 0.693 |
| 0.75 | 1.386 |
| 0.95 | 2.996 |

Verteilungsfunktion




Dichtefunktion



## 1.5.2 Charakteristische Größen von Verteilungen

- Die Gesamtinformation, die mit einer Wahrscheinlichkeitsverteilung gegeben wird (oder gegeben werden muss), ist häufig zu umfangreich, deshalb nutzt man abgeleitete Kenngrößen, die in praktischen Situationen gut zu nutzen sind. Dabei kann man bei den Kenngrößen im Allgemeinen Lageparameter und Streuungsparameter unterscheiden.
- Die am häufigsten genutzte Kenngröße ist der Erwartungswert $\mathbf{E}X$ einer Zufallsgröße $X$ (auch Mittelwert der Zufallsgröße genannt). Er ist ein Lageparameter, eine (nichtzufällige) reelle Zahl und beschreibt den Schwerpunkt der Wahrscheinlichkeitsmasse.

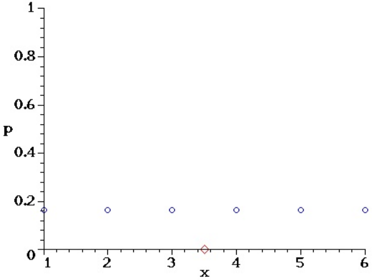
# Erwartungswert einer Zufallsgröße I

- **Definition:** Für eine diskrete Zufallsgröße $X$ mit möglichen Werten $x_1, x_2, \ldots$ und zugehörigen Wahrscheinlichkeiten $p_1 = P(X = x_1)$, $p_2 = P(X = x_2)$, ... wird der Erwartungswert definiert durch
$$\mathbf{E}X = \sum_i x_i p_i \,.$$
Für eine stetige Zufallsgröße $X$ mit Dichtefunktion $f_X$ wird der Erwartungswert definiert durch
$$\mathbf{E}X = \int_{-\infty}^{\infty} x \cdot f_X(x) \, \mathrm{d}x \,.$$

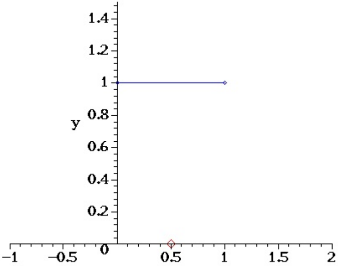
# Erwartungswert einer Zufallsgröße II

**Beispiele 1.1 und 1.2:**

$X$ Augenzahl beim Würfeln
Einzelwahrscheinlichkeiten
und Erwartungswert




$X$ gleichverteilt auf $[0, 1]$
Dichtefunktion
und Erwartungswert



# Erwartungswert einer Zufallsgröße III

- Es gelten folgende Rechenregeln für Erwartungswerte:
  Sind $X$ und $Y$ Zufallsgrößen und $a$ und $b$ reelle Zahlen, dann gelten
  $$\mathbf{E}(a + b \cdot X) = a + b \cdot \mathbf{E}X\,;$$
  $$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y\,.$$
  Dies sind die Linearitätseigenschaften der Erwartungswertbildung.

- Nicht jede Zufallsgröße besitzt einen Erwartungswert.

- Ist $g : \mathbb{R} \to \mathbb{R}$ eine (z.B. stetige) Funktion und $X$ eine Zufallsgröße, dann kann man den Erwartungswert der Zufallsgröße $Y = g(X)$ wie folgt berechnen:
  $$\mathbf{E}Y = \mathbf{E}g(X) = \sum_i g(x_i)p_i \qquad \text{für eine diskrete ZG} \quad X\,;$$
  $$\mathbf{E}Y = \mathbf{E}g(X) = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x \qquad \text{für eine stetige ZG} \quad X\,.$$

# Varianz einer Zufallsgröße

- Die wichtigste Kenngröße für die Variabilität von Zufallsgrößen ist die Varianz der Zufallsgröße, auch Streuung oder Dispersion genannt. Sie gibt die erwartete quadratische Abweichung der Zufallsgröße von ihrem Erwartungswert an.
- **Definition:** Die Varianz $\mathbf{Var}X$ der Zufallsgröße $X$ ist die nichtnegative reelle Zahl (falls sie existiert)

$$\mathbf{Var}X = \mathbf{E}\left(X - \mathbf{E}X\right)^2 = \begin{cases} \sum\limits_i \left(x_i - \mathbf{E}X\right)^2 p_i\,, & \text{diskrete ZG;} \\ \int\limits_{-\infty}^{\infty} \left(x - \mathbf{E}X\right)^2 f_X(x)\,\mathrm{d}x\,, & \text{stetige ZG.} \end{cases}$$

- Die Varianz lässt sich meistens bequemer mit Hilfe der Formel

$$\mathbf{Var}X = \mathbf{E}\left(X^2\right) - \left(\mathbf{E}X\right)^2$$

berechnen.

# Standardabweichung einer Zufallsgröße und Eigenschaften

- **Definition:** Die Standardabweichung $\sigma_X$ der Zufallsgröße $X$ ist die positive Quadratwurzel aus der Varianz der Zufallsgröße:

$$\sigma_X = \sqrt{\mathbf{Var}X}\,.$$

- Ist $a$ eine reelle Zahl und $X$ eine Zufallsgröße, dann gelten
  - $\mathbf{Var}(aX) = a^2\mathbf{Var}X$ ,
  - $\mathbf{Var}(a + X) = \mathbf{Var}X$ ,
  - $\sigma_{(aX)} = |a|\sigma_X$ ,
  - $\sigma_{(a+X)} = \sigma_X$ .
- Es gilt genau dann $\mathbf{Var}X = \sigma_X = 0$, wenn es eine reelle Zahl $x_0$ gibt, so dass $P(X = x_0) = 1$ gilt.

# Berechnung der Varianzen in den Beispielen 1.1 und 1.2

- $X$ Augenzahl beim Würfeln mit einem gerechten Würfel.

$$\mathbf{E}X^2 = \frac{1^2}{6} + \frac{2^2}{6} + \frac{3^2}{6} + \frac{4^2}{6} + \frac{5^2}{6} + \frac{6^2}{6} = \frac{91}{6}$$

$$\mathbf{Var}X = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} = 2.917.$$

- $X$ gleichmäßig verteilt auf dem Intervall $[0, 1]$.

$$\mathbf{E}X^2 = \int_0^1 x^2 \cdot 1 \, \mathrm{d}x = \frac{1}{3}$$

$$\mathbf{Var}X = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12} = 0.0833.$$

# Standardisierung und Variationskoeffizient

- **Definition:** Für eine Zufallsgröße $X$ mit endlicher Varianz wird die Standardisierung definiert durch

$$Z = \frac{X - \mathbf{E}X}{\sigma_X}\,.$$

  Dies ist eine mit X zusammenhängende Zufallsgröße, die den Erwartungswert 0 und eine Varianz von 1 besitzt.

- **Definition:** Für eine Zufallsgröße $X$ mit endlicher Varianz und $\mathbf{E}X \neq 0$ wird der Variationskoeffizient $\mathbf{V}_X$ definiert durch

$$\mathbf{V}_X = \frac{\sigma_X}{\mathbf{E}X}\,.$$

  Mit ihm wird die Streuung der möglichen Werte zum mittleren Wert (Erwartungswert) in Beziehung gesetzt, dadurch hilft er beim Vergleich der möglichen zufälligen Schwankungen der Werte von Zufallsvariablen.

# Kovarianz und Unkorreliertheit zweier Zufallsgrößen

- Für zwei Zufallsgrößen $X$ und $Y$ mit endlicher Varianz heißt die reelle Zahl

$$\mathbf{Cov}\,(X,Y) = \mathbf{E}\,((X - \mathbf{E}X)(Y - \mathbf{E}Y)) = \mathbf{E}(XY) - \mathbf{E}X \cdot \mathbf{E}Y$$

  die Kovarianz der beiden Zufallsgrößen. Sie ist ein Maß für die Stärke eines linearen Zusammenhangs zwischen $X$ und $Y$.

- Der Korrelationskoeffizient der Zufallsgrößen $X$ und $Y$ ist dann

$$\varrho_{X,Y} = \mathbf{Corr}\,(X,Y) = \frac{\mathbf{Cov}\,(X,Y)}{\sigma_X \sigma_Y}\,.$$

  Dieser Wert liegt immer zwischen -1 und 1. Im Fall $|\varrho_{X,Y}| = 1$ besteht ein vollständiger linearer Zusammenhang zwischen beiden Größen.

- Zwei Zufallsgrößen $X$ und $Y$ heißen unkorreliert, falls $\mathbf{Cov}\,(X,Y) = 0$ gilt.

# Unabhängigkeit von Zufallsgrößen und Varianz einer Summe von unabhängigen Zufallsgrößen

- **Definition:** Zwei Zufallsgrößen $X$ und $Y$ heißen stochastisch unabhängig, falls für beliebige reelle Zahlen $x, y$ gilt:

$$P\left(\{X < x\} \cap \{Y < y\}\right) = P(X < x) \cdot P(Y < y)\,.$$

- Sind zwei Zufallsgrößen $X$ und $Y$ mit endlichen Erwartungswerten stochastisch unabhängig, dann gilt $\mathbf{E}(X \cdot Y) = \mathbf{E}X \cdot \mathbf{E}Y$ . Damit sind $X$ und $Y$ auch unkorreliert.

- **Satz:** Sind zwei Zufallsgrößen $X$ und $Y$ stochastisch unabhängig (oder unkorreliert), dann gilt für deren Summe:

$$\mathbf{Var}(X + Y) = \mathbf{Var}X + \mathbf{Var}Y\,.$$

Diese Eigenschaft gilt aber nicht im Allgemeinen!

# Tschebyschew-Ungleichung

- Man kann mit Hilfe von Erwartungswerten und Varianzen auch Wahrscheinlichkeiten abschätzen. Dabei finden die folgenden Ungleichungen öfters Verwendung.

- **Satz:**
  Für eine Zufallsgröße $X$ mit $\mathbf{E}|X| < \infty$ gilt für beliebige $c > 0$

  $$P(|X| \geq c) \leq \frac{\mathbf{E}|X|}{c}\,.$$

  Ist die Varianz der Zufallsgröße $X$ endlich, d.h.

  $$\mathbf{Var}X = \mathbf{E}(X - \mathbf{E}X)^2 < \infty\,,$$

  dann gilt auch

  $$P(|X - \mathbf{E}X| \geq c) \leq \frac{\mathbf{Var}X}{c^2}\,.$$

# Illustration Tschebyschew-Ungleichung für eine Exponentialverteilung mit Parameter $\lambda = 1$

Vergleich exakte Wahrscheinlichkeiten (blau) und Abschätzungen aus Tschebyschew-Ungleichung (rot) : für $c > 1$ :

links: $P(X \geq c) = 1 - F_X(c) = \mathrm{e}^{-c} \;\leq\; \dfrac{\mathbf{E}|X|}{c} = \dfrac{1}{c}$ ;

rechts: $P(|X - \mathbf{E}X| \geq c) = P(X - 1 \geq c) = \mathrm{e}^{-c-1} \;\leq\; \dfrac{\mathbf{Var}X}{c^2} = \dfrac{1}{c^2}$ .