KATA LOGO Mathematik – Statistik – Wahrscheinlichkeitsverteilungen

x = reelle Zahl
X = Zufallsgröße

1. Diskrete Zufallsgrößen

Beschreibung durch: **a) Liste ihrer Merkmalswerte (x1,x2, ...)**
b) dazugehörige Wahrscheinlichkeiten

Ermittlung von Wahrscheinlichkeiten: **a) Formeln nutzen und rechnen**

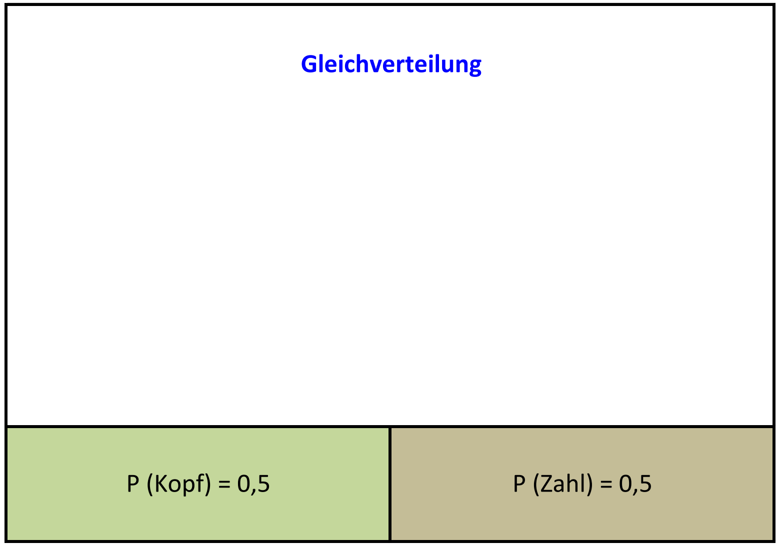
1. Nicht parametrische Verteilungen: Ohne Verteilungsannahme

2. Parametrische Verteilungen: Grundannahme: ZV kommt aus einer bestimmten WS-Verteilung

Variablen sind diskret verteilt	**Induktive Statistik (Inferenzstatistik = schließende Statistik)**	**Deskriptive Statistik**
	Schätz- und Testverfahren gibt die Wahrscheinlichkeitstheorie vor	**Wie kann man eine Verteilung eines Merkmals beschreiben?**
Binomialverteilung · Anwendung bei vielen Wiederholungen	*Erwartungswert*	*Mittelwert*
Poisson-Verteilung	*Theoretische Varianz*	*Empirische Varianz*
Hypergeometrische Verteilung · Ziehen ohne Zurücklegen	*Kumulative Verteilungsfunktion* · gibt die WS in einem bestimmten Bereich an · beschreibt Wahrscheinlichkeitsverteilung · kumuliert Zähldichte durch Summenbildung	*Kumulierte relative Häufigkeit*
Logarithmische Verteilung	*Wahrscheinlichkeitsfunktion* (Dichtefunktion = Zähldichte)	*Häufigkeitsverteilung*
Bernoulli-Verteilung		
Diskrete Gleichverteilung		

Variablen sind stetig verteilt	**Induktive Statistik (Inferenzstatistik = schließende Statistik)** **Schätz- und Testverfahren gibt die Wahrscheinlichkeitstheorie vor**
Normalverteilungen · Konfidenzschätzungen iso Hypothesentest - · Zentraler Grenzwertsatz bei unabhängigen Werten	*Dichtefunktion f(x)* · ist i. d. R. nicht bekannt · WS-Funktion von stetigen ZVen · WSen liegen in einem Intervall · empirische DF = relative Häufigkeit der Daten
Standardnormalverteilung N(0,1)	*Verteilungsfunktion* Stammfunktion F(X) · gibt die WS in einem bestimmten Bereich an · beschreibt Wahrscheinlichkeitsverteilung · kumuliert Zähldichte durch Summenbildung
χ^2-Verteilung (Stichprobenverteilung) · schätzt σ aus NV (auch μ und λ) FG	*Erwartungswert*
Studentsche t-Verteilung · gute Näherung für NV, aber schmäler · FG : n-1 bei Δ Dichtefunktion · schätzt μ, σ, λ (da man λ oft nicht hat)	*Theoretische Varianz*
F-Verteilung · mit 2 Parametern = FG: m1 und m2	
Exponentialverteilung E(λ) · Wartezeiten auf ... · Diskrete Variante: Geometrische Verteilg.	
F-Verteilung · durch 2 FG (m und n) festgelegt · über einem halbseitig ∞ Intervall · üblicherweise als [0,∞] angenommen	
Stetige Gleichverteilung	
Log-Normalverteilung	
Weibull-Verteilung	

1.1. Zufallsgrößen Typ A (alternativ)



Zufallsgrößen mit nur 2 möglichen Ereignissen bzw. Realisierungen.

Beispiele: Kopf oder Zahl, 1 oder 0, Urne mit nur blauen und roten Kugeln.

Beide Ereignisse sind gleichwahrscheinlich.

Verteilung: Gleichmäßig (50/50 oder 1/2 und 1/2)

WS-Funktion: **$P(X=t) = f_x(t) = 1/n$ für $t = xi$ ($i = 1, \ldots n$), sonst 0**

Verteilungs-Funktion: **$Fx(t) = P(X) \leq t) = 1/n \cdot | \{k : x_k \leq t\} |$**

EW: **$(n+1)/2$**

Var: **$(n^2-1)/2$**

1.2. Zufallsgrößen Typ D (4 verschiedene)

Beschreibung durch: **a) Liste ihrer Merkmalswerte (x1,x2, ...)**
b) dazugehörige Wahrscheinlichkeiten

Ermittlung von Wahrscheinlichkeiten: **a) Formeln nutzen und rechnen**

Gleichmäßig diskrete Verteilung

P (1) = 1/5	P (2) = 1/5	P (3) = 1/5	P (4) = 1/5	P (5) = 1/5

Zufallsgrößen als Liste mehrerer möglicher Ereignisse bzw. Realisierungen.

Beispiele: Würfel {1,2,3 4,5,6}

Alle Ereignisse sind gleichwahrscheinlich.

Verteilung: Gleichmäßig (im Würfel 1/6 pro Zahl)

WS-Funktion: **$P(X=t) = f_x(t) = 1/n$ für $t = xi$ $(i= 1, ...n)$, sonst 0**

Verteilungs-Funktion: **$Fx(t) = P(X) \leq t) = 1/n \cdot | \{k:x_k \leq t\} |$**

EW: $(n+1)/2$

Var: $(n^2-1)/2$

Beschreibung durch: **a) Liste ihrer Merkmalswerte (x1,x2, ...)**
b) dazugehörige Wahrscheinlichkeiten

Ermittlung von Wahrscheinlichkeiten: **a) Formeln nutzen und rechnen**

Binomiale Verteilung
Wenn sonst jeder zu 26% trifft, wie ist WS,
bei 11 Versuchen für genau 4 Treffer?

	x						x		
					x				
								x	

Beispiel: Der Treffer-Durchschnitt aller bisherigen Teilnehmer liegt bei 0,26. Wie hoch ist WS, bei 11 Versuchen genau 4 Treffer zu haben?

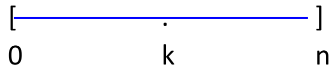
Zufallsexperimente mit "Wettkampf-Charakter" (mit Zurücklegen). Kleine Stichproben.
Faustregel: n/N ≤ 0,05. Setzung: p=M/N

WS für Eintreten des Einzelereignisses. Gezogene Elemente ändern P beim nächsten Zug wenig.

Gleiche Treffer-WS bei n-maligen unabhängigen Einzelversuchen (mit Zurücklegen = Bernoulli).

Verteilung: Zweiparametrig (A und A quer)

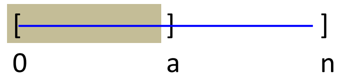
WS-Funktion: $P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$



Für große n erhält man über die Standardisierung der Binomialverteilung (setze z: = (X-µ)/σ) die Näherungsformel für die n → ∞ Gaußverteilung.

Verteilungsfunktionen:

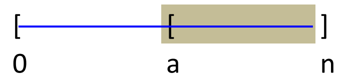
$P(X \leq a) = \sum_{k=0}^{a} \binom{n}{k} p^k \cdot (1-p)^{n-k} = F_x(a)$



EW: $n \cdot p$

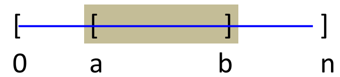
Var: $n \cdot p\,(1-p)$

$P(X \geq a) = \sum_{k=a}^{n} \binom{n}{k} p^k \cdot (1-p)^{n-k} = 1 - F_x(a-1)$



Großer Teich mit 10.000 Fischen. Nimmt man 5 Fische raus, macht es kein Unterschied für die Wahrscheinlichkeit. Ohne Zurücklegen.

$P(a \leq X \leq b) = \sum_{k=a}^{b} \binom{n}{k} p^k \cdot (1-p)^{n-k} = F_x(b)-F_x(a-1)$



Beschreibung durch: **a) Liste ihrer Merkmalswerte (x1,x2, ...)**
b) dazugehörige Wahrscheinlichkeiten

Ermittlung von Wahrscheinlichkeiten: **a) Formeln nutzen und rechnen**

Hypergeometrische Verteilung

Wenn ich aus 100 Teilen nur 3 zufällig ziehe, wie ist WS für Treffer bei 10 Ziehungen?

									defekt und versteckt
	defekt und versteckt								
				defekt und versteckt					
								defekt und versteckt	
		defekt und versteckt							

Großer Teich mit 300 Fischen. Nimmt man 20 Fische raus, macht es durchaus einen Unterschied für die Wahrscheinlichkeit.

Zufallsexperimente mit vergleichbarem "Entnahme-Charakter" (ohne Zurücklegen).

Beispiel: Anwendung bei Qualitätskontrollen. Maschine 100 N Teile, 5 k defekte.
WS für (k=3) aus 10er Stichprobe?

Bei Stichprobe (etwa $n/N < 0{,}05$) ähnliche Wahrscheinlichkeiten wie Binomialverteilung.
Binomialverteilung wird dann angewendet, da einfacher.

Da komplex, Variablen 1. k Treffer 2. K Defekte und 3. n Stichprobe hinschreiben.

Verteilung: Dreiparametrig.

WS-Funktion: $P(X=k) = \binom{K}{k} \cdot \binom{N-K}{n-k} / \binom{N}{n}$

Verteilungs-Funktion: $F_x = P(A \leq k) = \Sigma$ k=0 bis K $\binom{K}{k} \cdot \binom{N-K}{n-k} / \binom{N}{n}$

EW: $n \cdot (M/N)$

Var: $n \binom{M}{N} \cdot (1 - M/N) \cdot [(N\text{-}n) / N\text{-}1)]$

Punktschätzung für K (K^): ...

Punktschätzung für N (N^): ...

Beschreibung durch: **a) Liste ihrer Merkmalswerte (x1,x2, ...)**
b) dazugehörige Wahrscheinlichkeiten

Ermittlung von Wahrscheinlichkeiten: **a) Formeln nutzen und rechnen**

Poissonverteilung

Bei Ø 6 Personen/min (λ), wie ist WS für 9 Besuche **in der nächsten Minute**? Antwort: 6,88%

Normaler Wert pro Minute = λ

WS, dass es **in der nächsten Minute** genau 9 sind: P = 6,88%

Zufallsgrößen mit Realisierungen aus der Liste (0,1,2,...). Zufallsgröße, die mit zufälligen Ankünften in Verbindung steht.

Beispiel: Kaufhausbesuche gibt's λ [6 Personen pro Minute]. Bei Zählung im Minutentakt, wie groß ist WS, dass in nächster Minute k Personen das Kaufhaus betreten? (λ immer > 0).

Näherungsformel für Verteilung seltener Ereignisse. Für Phänomene, die innerhalb einer Einheit auftreten (8 Anrufe im Ø pro Minute).

WS-Funktion: $P(X=k) = (\lambda k/k!) \cdot e^{-\lambda}$

Verteilungs-Funktion: $Fx(n) = e^{-\lambda} \sum_{k=0}^{n} \lambda k/k!$

EW: λ

Var λ

2. Stetige Zufallsgrößen

2.1. Zufallsgrößen Typ S

Da Angabe der Merkmalswerte und deren WS unmöglich ist, gibt man für jede reelle Zahl x die WS an, daß die Realisierungen der Zufallsgröße X in dem links-offenen Intervall (-∞,x) liegt.

Beschreibung durch:

- a) Verteilungsfunktion
- b) Dichtefunktion

Ermittlung von Wahrscheinlichkeiten:

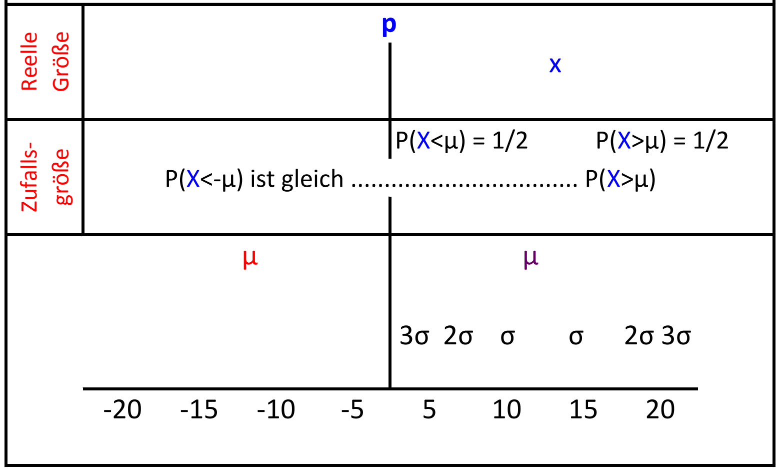
- a) Formeln nutzen und rechnen - geht nicht, da rechtstehendes Integral sich nicht auswerten lässt !!!
- b) Daher auf Tabellen zurückgreifen. Tabellen gibt es aber nur für die standardisierte NV.
- c) Daher entweder Excel oder R benutzen oder d)
- d) Umformen aus standardisierter Zufallsgröße $Z = (X-\mu)/\sigma$. Beispiel in e)
- e) Bei $\mu=4,8$ und $\sigma=3,9$ ausrechnen der WS $P(X<2)$:

  $P(X<x) = P\,((X-\mu)/\sigma<(x-\mu)/\sigma) \Rightarrow$

  $P(X<2) = P\,((X-4,8)/3,9<(2-4,8)/3,9) \Rightarrow P(Z<-0,72)$
- f) Schließlich in Tafel der Standard-NV schauen beim Wert -0,72, um WS zu finden.
- g) Da Tabelle erst bei 0 anfängt, umformen wegen Symmetrie: $P(Z < -\,0,72) = 1-P(Z< +\,0,72)$

Normalverteilung (Gauß)

Verteilungsfunktion ist Stammfunktion F(x)



WS einer Zufallsgröße (X), daß ihre Realisierung bei gegebener Zahl x in dem links-offenen Intervall (-∞,x) liegt.

Zufallsgrößen sind normalverteilt mit Parametern μ und σ.

Die meisten WS-Verteilungen lassen sich bei großer Stichprobe zur NV überleiten.

Gut für Prozesse in Naturwissenschaft und Einkommensverteilung, die in mehreren Faktoren unabhängig voneinander in verschiedene Richtungen wirken.

$P(\mu-3\sigma \le X < \mu+3\sigma) = 0{,}997$
$P(\mu-2\sigma \le X < \mu+2\sigma) = 0{,}945$
$P(\mu-\sigma \le X < \mu+\sigma) = 0{,}683$

WS-Funktion (Dichtefunktion): $f_x(t) = 1 / \sigma\sqrt{2\pi} \cdot e \wedge -0{,}5\, [(t-\mu)/\sigma)]^2$

Verteilungs-Funktion (komplex): $F_x(t) = 1 / \sigma\sqrt{2\pi} \int_{-\infty}^{t} e \wedge -0{,}5\, [(s-\mu)/\sigma)]^2\, ds$

Verteilungs-Funktion (vereinfacht): $\Phi_{\mu,\sigma}(x) = \Phi\,(X-\mu / \sigma)$ mit $x \in R$

$P(X \le b) = F(b) = \Phi\,(b-\mu / \sigma)$

$P(X \ge a) = 1 - F(a) = 1 - \Phi\,(a-\mu / \sigma) = 1 - \Phi$ (z. B. 2) $= \Phi$ (z. B. -2)

$P(a \le X \le b) = F(b) - F(a) = \Phi\,(b-\mu / \sigma) - \Phi\,(a-\mu / \sigma)$

mit $\Phi_{0,1} = \Phi$ und $\Phi(0) = 1/2$

EW: μ

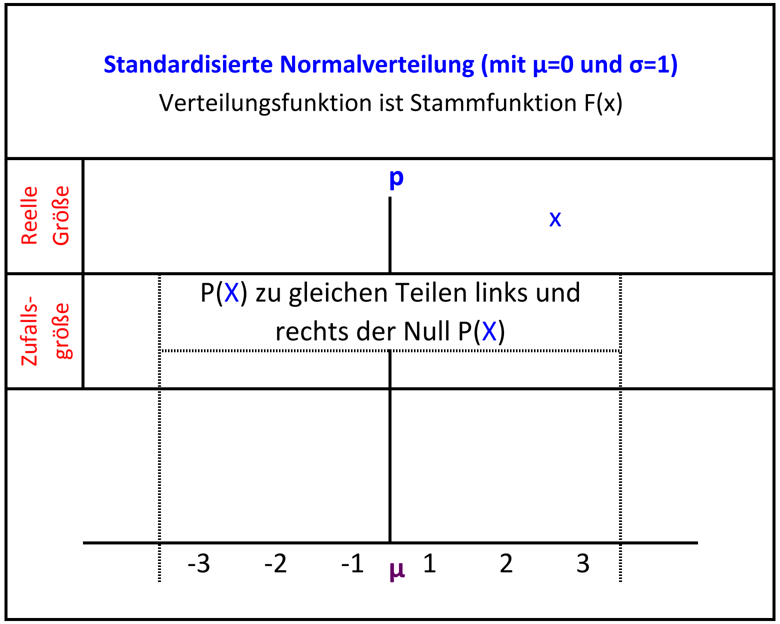
Var: σ

Median: μ

WS-Funktion (Dichtefunktion) ff:

EW: $\mu = \mu_x = \int_{-\infty}^{\infty} t \cdot f(t)\, dt$

Var: $\sigma^2 = \sigma x^2 = \int_{-\infty}^{\infty} (t - \mu)^2 \cdot f(t)\, dt$



WS einer Zufallsgröße (X), daß ihre Realisierung bei gegebener Zahl x in dem links-offenen Intervall (-∞,x) liegt.

WS-Funktion (Dichtefunktion) ff:

EW: $\mu = \mu_x = \int_{-\infty}^{\infty} t \cdot f(t)\,dt$

Var: $\sigma^2 = \sigma x^2 = \int_{-\infty}^{\infty} (t - \mu)^2 \cdot f(t)\,dt$

Zufallsgrößen sind normalverteilt mit Parametern μ=0 und σ=1. Realisierungen liegen zu gleichen Teilen links und rechts der Null - fast ausschließlich im Intervall -3 und +3.

Beispiel:

WS für die Standardnormalverteilung [Φ(z)] aus der Tabelle holen. Dabei ist [Φ(-z)] = 1 - [Φ(z)].

P(-3 ≤ X < 3) = 0,9973
P(-2 ≤ X < 2) = 0,9545
P(-1 ≤ X < 1) = 0,6827

WS-Funktion (Dichtefunktion): **$f_x(x) = 1 / \sqrt{2\pi} \cdot e \wedge -0{,}5\, x^2$**

Verteilungs-Funktion: **$F_x(x)$ oder $\Phi(x) = 1 / \sqrt{2\pi} \int_{-\infty}^{x} e \wedge -0{,}5\, s^2\, ds$**

P(X ≤ b) = F(b) = Φ (b)

P(X ≥ a) = 1 - F(a) = 1 - Φ (a) = 1 - Φ (z. B. 2) = Φ (z. B. -2)

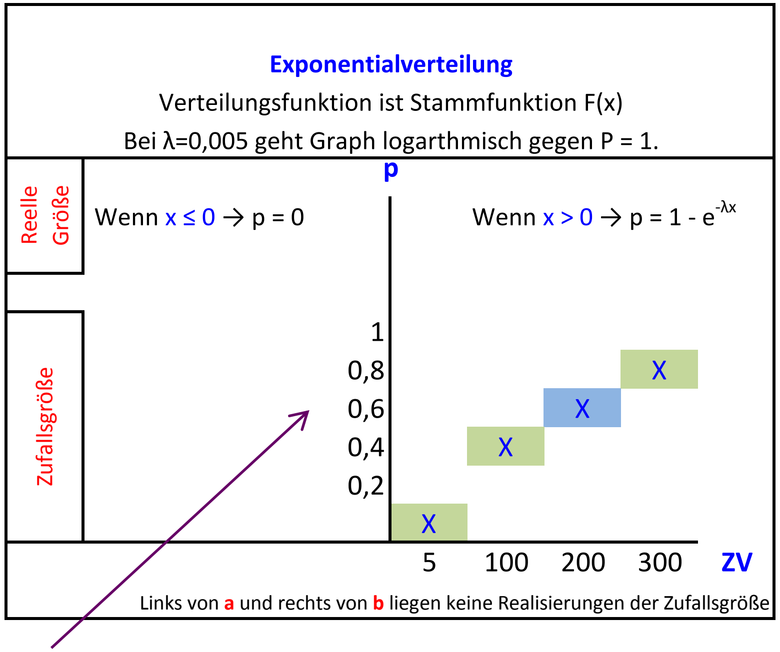
P(a ≤ X ≤ b) = F(b) - F(a) = Φ (b) - Φ (a)

mit **$\Phi_{0,1} = \Phi$** und **Φ (0) = 1/2**

EW: 0

Var: 1

Median: 0



WS einer Zufallsgröße (X), daß ihre Realisierung bei gegebener Zahl x in dem links-offenen Intervall (-∞,x) liegt.

Zufallsgröße "Zeitdauer der Funktionsfähigkeit eines Weckers in Tagen" mit Ausfallrate **λ**.

Gibt die Dauer von zufälligen Zeitintervallen an. Mittlere Reichweite oder Lebensdauer (z. B. von Atomen beim radioaktiven Zerfall); Lebensdauer von Maschinen, wenn Alterungserscheinungen nicht betrachtet werden müssen; Zeit zwischen zwei Anrufen; KFZ-Haftpflicht.

Beispiel: Unabhängig von ihrem Alter ausfallende Wecker pro Tag: 5 ‰. Bei Ausfallrate λ= 0,005 ist Ø Zeitdauer, bis ein Wecker ausfällt, 1/**λ** = 200 Tage.

Auf Lebewesen darf sie nicht angewendet werden, sonst wäre z. B. die WS, dass ein Achtzigjähriger noch weitere 50 Jahre lebt, genauso hoch wie die, dass ein Neugeborener das fünfzigste Lebensjahr erreicht.

Wird durch einzigen Parameter **λ** bestimmt. Er besitzt den Charakter einer Ereignisrate (1/**λ**). Wert von **λ** sind 0 (asymptotischer Graph) und ∞ (exponentialer Graph).

Anwendbar, wenn die poissonschen Annahmen erfüllt sind.

WS-Funktion (Dichtefunktion): $f_x(x) = \lambda e^{-\lambda x}$ **für** **x ≥ 0, sonst 0 für x < 0**

Verteilungs-Funktion: $F_x(x) = 1-e^{-\lambda x}$ **x ≥ 0, sonst 0 für x < 0**

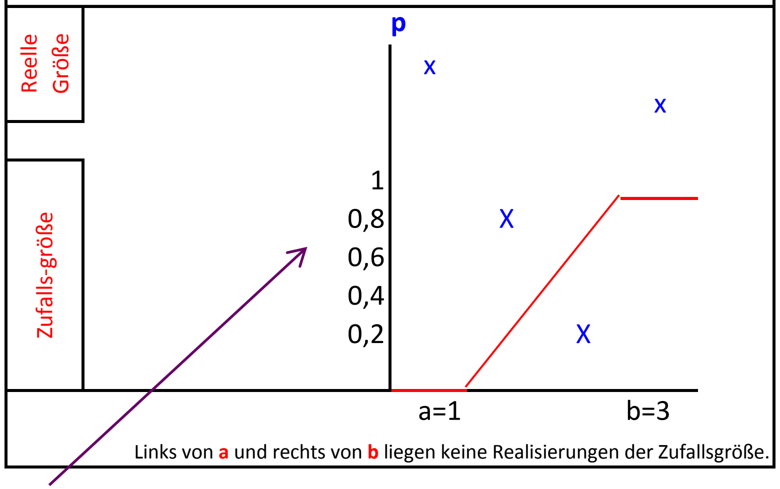
$1-F(X) = e^{-\lambda x}$:Komplement der Verteilungsfunktion ist die Überlebenswahrscheinlichkeit bei sogenannten ermüdungsfreien Systemen. Ausfallrate ist hierbei **λ**. Wahrscheinlichkeit, dass ein Wecker höchstens (noch) 20 Tage hält, ist $1 - e^{-0{,}005 \cdot 20\text{ Tage}} = 0{,}0952$, d.h. nach 20 Tagen sind durchschnittlich ca. 10 % der Wecker ausgefallen. Entsprechend ist der Anteil der Wecker, die mindestens 180 Tage aushalten: $1 - (1 - e^{-0{,}005 \cdot 180\text{ Tage}} = 1- 0{,}5934 = 0{,}4066$.

EW: 1/λ

Var 1/λ² und SA: 1/λ und Median: ln2/λ

Stetige Gleichverteilung

Verteilungsfunktion ist Stammfunktion F(x) ist monoton steigend.



WS einer Zufallsgröße (X), daß ihre Realisierung bei gegebener Zahl x zwischen a und b liegt.

Zufallsgrößen sind gleichmäßig verteilt - mit den Parametern **a** und **b**.

Rechteckverteilung, Uniformverteilung

Über einem endlichen Intervall [a,b], im einfachsten Fall [0,1]. Gutes Modell also, wenn X nur Werte in [a; b] annehmen kann und mit gleicher WS in alle gleich großen Teilbereiche fällt. Hat konstante Wahrscheinlichkeitsdichte auf einem Intervall (a,b): Das bedeutet: Alle Teilintervalle gleicher Länge besitzen dieselbe Wahrscheinlichkeit.

Beispiel: Shuttlebus fährt alle 20 min. Ein Fahrgast (der nicht weiß , zu welchen Uhrzeiten der Bus fährt) kommt zur Haltestelle. Die ZV für seine Wartezeit ist auf [0; 20]-gleichverteilt.

Häufig wird a = 0 und b = 1 angenommen. Dann ist F(x) = x auf dem Intervall [0,1].

$P(X<x) = 0$ $\quad x < a$
$P(X<x) = x-a/b-a$ $\quad a \leq x < b$
$P(X<x) = 1$ $\quad x > b$