5. stetige Verteilungen

Diskrete Stochastik

Prof. Dr. Andreas Vogt

Fachhochschule Nordwestschweiz

stetige Verteilungen

Eine stetige Zufallsvariable hat einen kontinuierlichen Wertebereich, bestehend aus einem Intervall oder ganz $\mathbb{R}$.

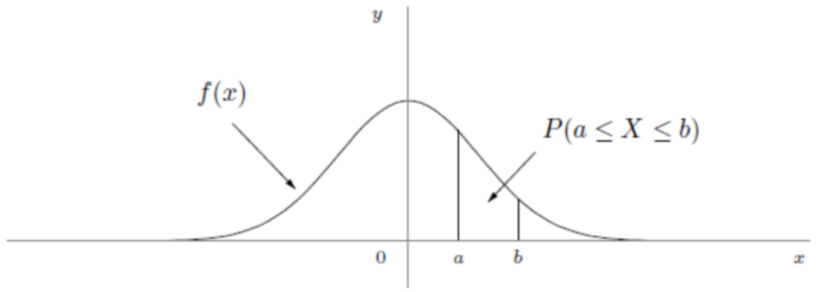
Etwa:
- Körpergrösse von erwachsenen Personen
- Lebensdauer einer Glühbirne

Die Wahrscheinlichkeit, dass eine stetige Zufallsvariable X einen exakten Wert x annimmt, z.B. dass die Körpergrösse einer Person gleich 175.17243 cm ist, ist gleich 0.

Sinnvoll sind Wahrscheinlichkeiten, dass X einen Wert in einem Intervall $[a, b]$ animmt.

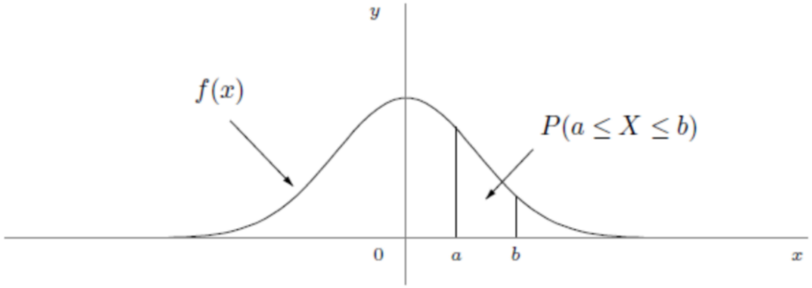
Diese Wahrscheinlichkeiten werden durch die Dichte $f(x)$ der Zufallsvariablen beschrieben:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$



stetige Verteilungen

- Die Dichtefunktion erfüllt $f(x) \geq 0$ für alle x, da Wahrscheinlichkeiten nicht negativ sein können.



- Die Gesamtfläche unter der Dichtefunktion muss gleich 1 sein:

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

- Abgeschlossene und offene Intervalle haben dieselben Wahrscheinlichkeiten:

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

Achtung: $f(x)$ ist nicht die Wahrscheinlichkeit für x.

Die Wahrscheinlichkeit für ein Intervall ist die Fläche unterhalb von $f(x)$

die stetige Gleichverteilung

Ein Dartpfeil wird zufällig auf das Intervall $[0,1]$ geworfen.
(Es wird nicht auf die Mitte gezielt.)

X gebe die Position des Darts an.

Was ist ein sinnvolles Modell?

Für $0 \leq a < b \leq 1$ sollte sein: $P(a \leq X \leq b) = b - a$.

Wie sieht also die Dichte aus?

$$P(a \leq X \leq b) = \int_a^b f(x)dx = b - a \text{ ist erfüllt für } f(x) = 1,\ x \in [0,1].$$

X heisst stetig gleichverteilt auf dem Intervall $[0,1]$.

Schreibweise: $X \sim U[0,1]$.

Für ein solches X gilt etwa: $P(X \in [0.3, 0.6]) = 0.3$

Erwartungswert, Varianz, Standardabweichung

Definition:

Es sei X ein Zufallsvariable mit Dichte f. Dann heisst (im Falle der Existenz)

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx \quad \text{Erwartungswert von } X$$

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x)dx \quad \text{Varianz von } X$$

$$= E(X^2) - E(X)^2$$

$$\sigma(X) = \sqrt{V(X)} \quad \text{Standardabweichung von } X$$

Ferner heisst $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$ Verteilungsfunktion von X.

Eigenschaften der Verteilungsfunktion:

- $\boxed{P(a \leq X \leq b) = F(b) - F(a)}$
- $\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1$
- $F'(x) = f(x)$ (falls f stetig)

die stetige Gleichverteilung

Definition:

Die stetige Gleichverteilung auf dem Intervall $[s,t]$ ist gegeben durch die konstante Dichte $f(x) = \frac{1}{t-s}$ für $s \le x \le t$.

Schreibweise: $X \sim U[s,t]$

$E(X) = \frac{s+t}{2}$ $\qquad E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \frac{1}{t-s}\int_s^t x dx = \frac{1}{t-s}\frac{t^2-s^2}{2} = \frac{s+t}{2}$

$V(X) = \frac{1}{12}(t-s)^2$

Matlab-Funktionen:

Dichte: $\texttt{unifpdf}(x,s,t)$

Verteilungsfunktion: $\texttt{unifcdf}(x,s,t)$

Beispiel:

Person A kommt zu einem zufälligen Zeitpunkt zum Bahnhof. Der Zug fährt stündlich. Wie hoch ist die Wahrscheinlichkeit, dass man höchstens 10 Minuten wartet? Was ist die erwartete Wartezeit?

Modellierung: X: Wartezeit

Dann ist $X \sim U[0,60]$ eine sinnvolle Modellierung.

Wir erhalten: $P(X \le 10) = \int_0^{10} \frac{1}{60} dx = \frac{10}{60} = \frac{1}{6}$.

$E(X) = 30$

Die erwartete Wartezeit beträgt also 30 Minuten.

die Normalverteilung

Die Normalverteilung von C.F.Gauss ist die wichtigste stetige Verteilung.

30. April 1777 – 23. Februar 1855

Sie tritt immer dann auf, wenn eine Zufallsvariable eine Summe von vielen unabhängigen Summanden ist. Dies ist der Inhalt des zentralen Grenzwertsatzes (siehe später).

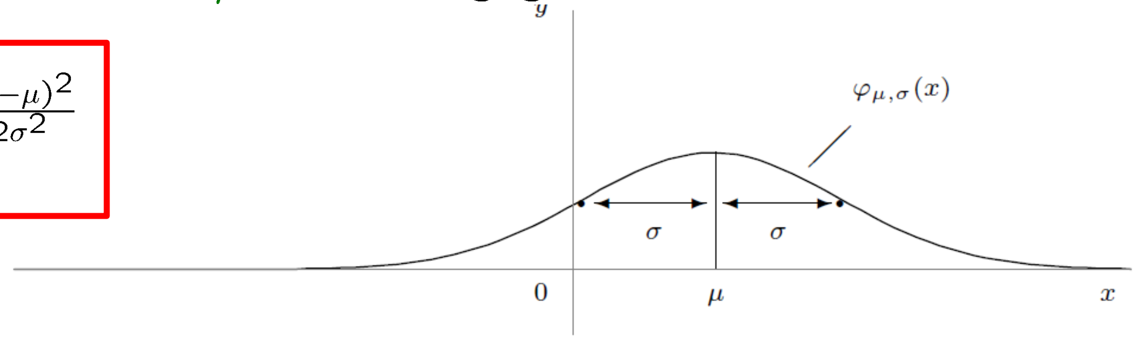
Da in der Natur viele Grössen, z.B. die Körpergrösse von Personen, Summen von vielen Einflüssen sind, tritt die Normalverteilung häufig in Erscheinung.

die Normalverteilung

Definition:

Die Normalverteilung mit Parametern μ und σ ist gegeben durch die Dichte

$$\boxed{\varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$



Schreibweise: $X \sim \mathcal{N}(\mu, \sigma)$

$E(X) = \mu$

$V(X) = \sigma^2$

Matlab-Funktionen:

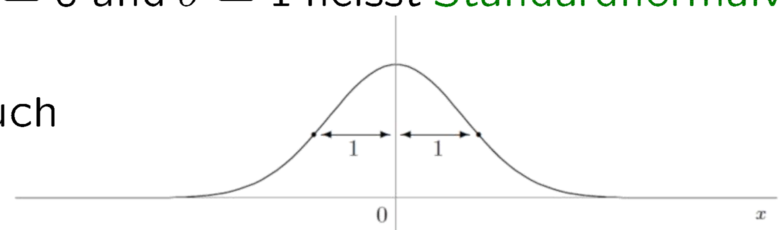
Dichte: $\texttt{normpdf}(x, \mu, \sigma)$

Verteilungsfunktion: $\texttt{normcdf}(x, \mu, \sigma)$

Es gibt keinen “geschlossenen Ausdruck” für die Verteilungsfunktion, man kann sie nur mit Computern vernünftig berechnen.

Die Normalverteilung mit den Parametern $\mu = 0$ und $\sigma = 1$ heisst Standardnormalverteilung.

Die Dichte $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ nennt man auch Gauss’sche Glockenkurve.



die Normalverteilung

Beispiel:

Der Intelligenzquotient (IQ) ist normalverteilt und so festgelegt, dass $\mu = 100$ und $\sigma = 15$. Wie hoch ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person einen IQ (a) zwischen 90 und 110 und (b) grösser als 150 hat?

Es sei X der IQ der Person.

Dann ist $X \sim \mathcal{N}(100, 15)$.

Wir erhalten:

$$P(90 \leq X \leq 110) = normcdf(110, 100, 15) - normcdf(90, 100, 15) \approx 50\%$$

$$P(X \geq 150) = 1 - P(X < 150) = 1 - normcdf(150, 100, 15) \approx 0.04\%$$

Serie 5, A.1 – A.7

Standardisierung

Es sei $X \sim \mathcal{N}(\mu, \sigma)$. Dann gilt:

- $X - \mu \sim \mathcal{N}(0, \sigma)$.
- $Z \sim \mathcal{N}(0, 1)$ mit $Z = \frac{X-\mu}{\sigma}$.

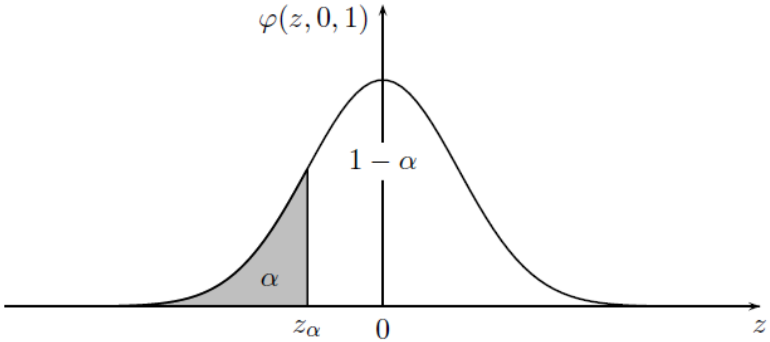
  Z heisst Standardisierung von X.

Oft (insbesondere bei statistischen Hypothesentests) hat man es mit folgender Fragestellung zu tun:

Gegeben ist ein $\alpha \in (0,1)$. Für welchen Wert z_α gilt $P(X \leq z_\alpha) = \alpha$?

Definition:

So ein z_α heisst α-Quantil (oder Perzentil).



Matlab Funktion: $norminv(\alpha, \mu, \sigma)$.

Beispiel:

Der Intelligenzquotient (IQ) ist normalverteilt und so festgelegt, dass $\mu = 100$ und $\sigma = 15$. Eine gewisse Schulform ist für die 5% der Bevölkerung gedacht, die den tiefsten IQ haben. Ab welchem IQ sollte man auf diese Schule gehen?

Für $X \sim \mathcal{N}(100, 15)$ ist also eine Zahl z gesucht mit $P(X \leq z) = 5\%$, also das 5%-Quantil von $\mathcal{N}(100, 15)$.

`norminv`$(0.05, 100, 15)$ in Matlab liefert 75.33.

die Normalverteilung

Sigma-Regeln

Für $X \sim \mathcal{N}(\mu, \sigma)$ gilt

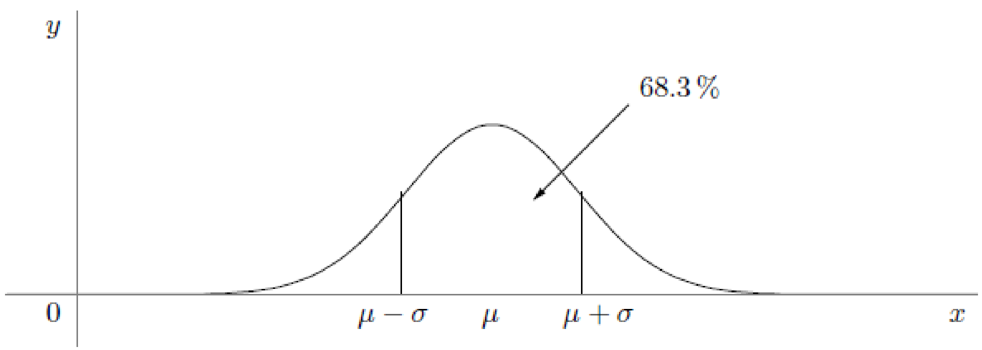
$P(|X - \mu| \leq \sigma) \approx 68.3\%$ 1-Sigma-Regel

$P(|X - \mu| \leq 2\sigma) \approx 95.5\%$ 2-Sigma-Regel

$P(|X - \mu| \leq 3\sigma) \approx 99.7\%$ 3-Sigma-Regel

Die erste Sigma-Regel besagt, dass ein Wert einer normalverteilten Zufallsvariable mit der Wahrscheinlichkeit 68% maximal um $\pm\sigma$ vom Erwartungswert μ abweicht.

Für Stichproben heisst dies, dass 68% der Werte im Intervall $\mu \pm \sigma$ zu erwarten sind.



Lernkontrolle

Die Grösse der Bäume von Bauer S. Claus ist normalverteilt. Die Bäume sind im Mittel 180cm gross bei einer Standardabweichung von 10cm.

1. Mit welcher Wahrscheinlichkeit ist ein Baum kleiner als 190cm?

2. Mit welcher Wahrscheinlichkeit ist ein Baum grösser als 170cm?

3. Wir kaufen 5 Bäume. Mit welcher Wahrscheinlichkeit sind davon mindestens zwei Bäume mindestens 170cm gross?

4. Ab wann zählt ein Baum zu den 10% grössten Bäumen?

die Exponential-Verteilung

An einer schwach befahrenen Landstrasse werden vorbeifahrende Fahrzeuge beobachtet, im Schnitt λ pro Minute

Es sei:

T : Zeit [Minuten] zwischen zwei aufeinanderfolgenden Fahrzeug-Ankünften

X : Anzahl Fahrzeug-Ankünfte in einer bestimmten Minute

Wir hatten uns schon überlegt, dass $X \sim Poi(\lambda)$ ein sinnvolles Modell ist.

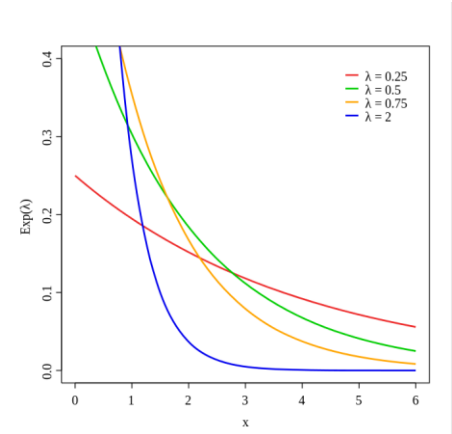
Für die Zufallsvariable X_t, welche die Ankunftsanzahlen im Zeitintervall t Minuten angibt, gilt also $\boxed{X_t \sim Poi(\lambda t)}$

Wie ist dann T verteilt?

$T > t$ heisst, dass in t Minuten niemand ankommt, also $X_t = 0$

$$P(T > t) = P(X_t = 0) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t},$$

also $P(T \leq t) = 1 - e^{-\lambda t}$.



Definition:

Eine Zufallsvariable mit Verteilungsfunktion $F(t) = 1 - e^{-\lambda t}$, also mit der Dichte $F'(t) = \lambda e^{-\lambda t}$ heisst exponentiell verteilt mit Parameter λ.

Schreibweise: $X \sim Exp(\lambda)$ $\quad E(X) = \frac{1}{\lambda}$ $\quad V(X) = \frac{1}{\lambda^2}$

Matlab-Funktionen:

Dichte: $\texttt{exppdf}(x, 1/\lambda)$

Verteilungsfunktion: $\texttt{expcdf}(x, 1/\lambda)$

die Exponential-Verteilung

Die Exponentialverteilung beschreibt zufällige Lebensdauern von Geräten oder Wartezeiten auf zufällige Ereignisse.

– Lebensdauer einer Glühbirne

– Wartezeit auf die nächste Vorbeifahrt eines Fahrzeuges an einer Landstrasse

– Wartezeit auf nächstes Erdbeben

– Wartezeit auf nächsten Anruf in einer Telefonzentrale

Wir hatten schon gesehen, dass die Zwischenankunftszeiten bei einer Poisson-verteilten Zufallsvariable exponentiell verteilt sind.

Umgekehrt kann man auch zeigen, dass aus exponentiell verteilten Zwischenankunftszeiten mit Parameter λ Possion-verteilte Ankunftsanzahlen mit Parameter λ folgen.

$$\boxed{\text{Also: Zwischenankunftszeit} \sim Exp(\lambda) \quad \Leftrightarrow \quad \text{Anzahl} \sim Poi(\lambda)}$$

die Exponential-Verteilung

Beispiel:

In ein Geschäft kommen pro Stunde im Schnitt 20 Kunden.

(1) Wie hoch ist die Wahrscheinlichkeit, dass mehr als 30 Kunden in einer Stunde kommen?

(2) Wie hoch ist die Wahrscheinlichkeit, dass man weniger als 5 Minuten auf den ersten Kunden warten muss?

(1) X zähle die Kunden pro Stunde. Dann ist $X \sim Poi(20)$ sinnvoll.

Wir erhalten $P(X > 30) = \boxed{1 - \texttt{poisscdf(30,20)} = 0.0135}$

(2) T beschreibe die Wartezeit in Stunden. Dann folgt $T \sim Exp(20)$ aus $X \sim Poi(20)$.

$$\text{Wir erhalten } P\left(T \leq \frac{1}{12}\right) = \int_0^{\frac{1}{12}} 20 \cdot e^{-20t} dt$$

$$= \texttt{expcdf}(1/12, 1/20) = 0.8111$$

die Exponential-Verteilung

Die Exponentialverteilung hat eine bemerkenswerte Eigenschaft, sie hat kein Gedächtnis. Das nennt man auch die *No Memory Property*.

Wenn ein Gerät mit einer exponentiell verteilten Lebensdauer X während t Stunden gelaufen ist, so gilt: die Wahrscheinlichkeit, dass es weitere h Stunden läuft ist gleich gross wie die Wahrscheinlichkeit, dass ein neues Gerät die ersten h Stunden läuft.

$$P(X \geq t+h \mid X \geq t) = \frac{P(X \geq t+h, X \geq t)}{P(X \geq t)} = \frac{P(X \geq t+h)}{P(X \geq t)} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h} = P(X \geq h).$$

Bei Zufallsvariablen $X_1, X_2, \ldots$ schreibt man $\{X_1 \in A_1, X_2 \in A_2, \ldots\}$ für das Ereignis, dass $X_1 \in A_1$ **und** $X_2 \in A_2$ **und** $\ldots$ ist.

Serie 5, komplett