5 Spezielle Verteilungen

5.1 Diskrete und stetige Zufallsvariablen

Repetition
Ein *Zufallsexperiment* ist ein Vorgang, bei dem folgende Bedingungen erfüllt sind:
- Der Vorgang lässt sich unter den gleichen äusseren Bedingungen beliebig oft wiederholen.
- Es sind mehrere sich gegenseitig ausschliessende Ergebnisse möglich.
- Das Ergebnis lässt sich nicht mit Sicherheit voraussagen, sondern ist zufallsbedingt.

Die möglichen, sich gegenseitig ausschliessenden Ergebnisse $\omega_1, \omega_2, \omega_3 \ldots$ des Zufallsexperiments werden zur Menge Ω, dem *Ergebnisraum* zusammengefasst: $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$.
Eine *Zufallsvariable* X ist eine Funktion, die jedem Ergebnis aus Ω eine reelle Zahl zuweist.

Wir unterscheiden zwei verschiedene Typen von Zufallsvariablen:

Definition
Bei einer *diskreten* Zufallsvariable gibt es immer Lücken zwischen den Werten, die die Zufallsvariable annehmen kann; in der Regel *zählt* eine diskrete Zufallsvariable etwas und kann entsprechend nur ganzzahlige Werte annehmen.
Eine *stetige* Zufallsvariable hat hingegen ein kontinuierliches Spektrum von möglichen Werten; oft erhält man diese Werte durch *Messung*.

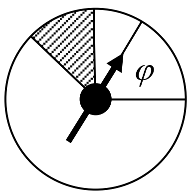
Beispiele

Zufallsvariable	diskret	stetig
Anzahl Studierende, die an einem bestimmten Tag anwesend sind	x	
Anzahl Tore in einem bestimmten Fussballspiel	x	
Temperatur an einem bestimmten Ort zu einer bestimmten Zeit		x
Höhe eines Heissluftballons zu einem bestimmten Zeitpunkt		x
Anzahl defekte Maschinenteile in einer Stichprobe	x	
Lebensdauer eines zufällig ausgewählten elektronischen Bauteils		x
Jahrgang eines zufällig ausgewählten Studenten	x	

In Kapitel 4 haben wir ausschliesslich diskrete Zufallsvariablen betrachtet. Wir zeigen nun, inwiefern sich stetige Zufallsvariablen von diskreten Zufallsvariablen unterscheiden.

Die *Dichtefunktion* (kurz *PMF*) f einer *diskreten* Zufallsvariablen X weist jeder reellen Zahl x die Wahrscheinlichkeit zu, mit der X den Wert x annimmt: $f(x) = P(X = x)$. Wie gross ist bei einer **stetigen** Zufallsvariable X die Wahrscheinlichkeit, dass sie einen bestimmten Wert x annimmt? Dazu ein Beispiel:

Beispiel
Wir betrachten das abgebildete Glücksrad. Die Zufallsvariable X beschreibt den Winkel φ zwischen der Horizontalen und dem Zeiger.



Es gilt: $P(0 \leq X < 2\pi) = 1$

Die Wahrscheinlichkeit, dass der Zeiger in einem bestimmten Sektor stehen bleibt, ist proportional zum Öffnungswinkel des Sektors, z.B.

$$P\left(\frac{\pi}{2} \le X < \frac{3\pi}{4}\right) = \frac{\frac{\pi}{4}}{2\pi} = \frac{1}{8}$$

Entsprechend gilt: $P(X = x) = 0$, denn der Öffnungswinkel ist hier 0! Und das, obwohl das Ergebnis $X = x$ nicht unmöglich ist…

Um trotzdem eine sinnvolle Dichtefunktion definieren zu können, müssen wir den Umweg über die CDF machen:

Definition
Die *kumulative Verteilungsfunktion* (kurz *CDF*) F ist *für diskrete und für stetige* Zufallsvariablen folgendermassen definiert: $F(x) = P(X \le x)$.

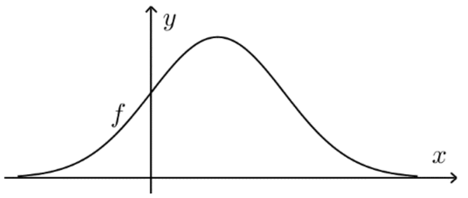
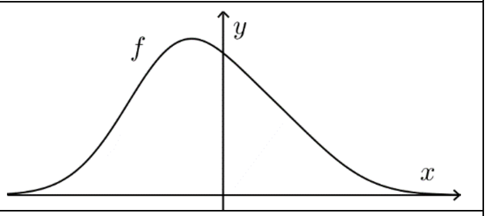
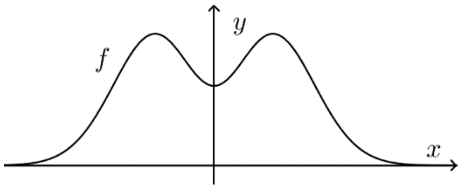
Definition
Bei einer *stetigen* Zufallsvariablen X lässt sich die Verteilungsfunktion als Integral einer Funktion f darstellen:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du$$

Der Integrand f heisst *Dichtefunktion* oder *PDF* der stetigen Zufallsvariable X .

Es gilt immer:	$f(x) \ge 0$ für alle $x \in \mathbb{R}$	$\int_{-\infty}^{\infty} f(u)du = 1$

Berechnung von Intervallwahrscheinlichkeiten

$P(X \le x) =$ $F(x) = \int_{-\infty}^{x} f(u)du$ x und $P(X \le x)$ einzeichnen.	
$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a) = \int_{-\infty}^{b} f(u)du - \int_{-\infty}^{a} f(u)du = \int_{a}^{b} f(u)du$ Erst Rechnung bis und mit vorletzter Ausdruck, dann Zeichnung.	
$P(X > x) = 1 - P(X \le x) = 1 - F(x) = \int_{x}^{\infty} f(u)du$	

Stetige Zufallsvariablen können ein kontinuierliches Spektrum von Werten annehmen. Deshalb müssen wir bei der Berechnung von Erwartungswert und Varianz Integrale statt Summen verwenden.

Definition
Gegeben ist eine *stetige* Zufallsvariable X mit der Dichtefunktion $f(x)$.
Dann ist der *Erwartungswert* von X definiert durch:

$$\mu = E(X) = \int_{-\infty}^{\infty} f(x) \cdot x \, dx$$

Die *Varianz* von X ist definiert durch:

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} f(x) \cdot (x - E(X))^2 \, dx$$

Die *Standardabweichung* ist die Quadratwurzel aus der Varianz: $\sigma = \sqrt{V(X)}$

Gegenüberstellung von diskreten und stetigen Zufallsvariablen

	diskrete Zufallsvariablen	**stetige Zufallsvariablen**
Dichtefunktion / PMF bzw. PDF	$f(x) = P(X = x)$	$f(x) = F'(x) \quad \neq \quad P(X = x)$!!!
Kumulative Verteilungsfunktion/ CDF	$F(x) = P(X \leq x) = \sum_{x \leq X} f(x)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u) du$
Wahrscheinlichkeiten	$P(a \leq X \leq b) = \sum_{a \leq x \leq b} f(x)$	$P(a \leq X \leq b) = \int_{a}^{b} f(x) dx$
Graphische Darstellung von f	Stabdiagramm	Graph
Erwartungswert	$E(X) = \sum_{x \in \mathbb{R}} f(x) \cdot x$	$E(X) = \int_{-\infty}^{\infty} f(x) \cdot x \, dx$
Varianz	$V(X) = \sum_{x \in \mathbb{R}} f(x) \cdot (x - E(X))^2$	$V(X) = \int_{-\infty}^{\infty} f(x) \cdot (x - E(X))^2 \, dx$

Für diskrete und stetige Zufallsvariablen X und Y gelten die folgenden Regeln:

(1) **Linearität des Erwartungswertes**:

$E(X + Y) = E(X) + E(Y)$ und $E(\alpha X) = \alpha E(X)$ mit $\alpha \in \mathbb{R}$.

(2) **Verschiebungssatz für die Varianz**: $V(X) = E(X^2) - (E(X))^2$
mit $E(X^2) = \int_{-\infty}^{\infty} f(x) \cdot x^2 \, dx$

(3) $V(\alpha X + \beta) = \alpha^2 \cdot V(X)$ mit $\alpha, \beta \in \mathbb{R}$.

(4) Sind X und Y unkorreliert ($COV(X, Y) = 0$), so gilt: $V(X + Y) = V(X) + V(Y)$.

Es gibt Typen von Verteilungen, die immer wieder auftreten. Da lohnt es sich, die grundlegenden Überlegungen einmal allgemein zu machen. Parameter erlauben es einem, die Verteilung der konkreten Aufgabenstellung anzupassen.
In den folgenden Abschnitten werden Sie einige der am meisten genutzten Verteilungen kennenlernen.

5.2 Diskrete Verteilungen

Die wichtigsten diskreten Verteilungen der Stochastik sind die *hypergeometrische Verteilung*, die *Binomialverteilung* (als Vorstufe dazu die *Bernoulliverteilung*) und die *Poisson Verteilung*. Diese Verteilungen korrespondieren mit den elementaren stochastischen Modellen für die drei Zufallsexperimente, dem Ziehen einer Stichprobe ohne Zurücklegen, dem Ziehen einer Stichprobe mit Zurücklegen und dem Auftreten voneinander unabhängiger gleichartiger Ereignisse in einem bestimmten Zeitfenster.

5.2.1 Hypergeometrische Verteilung

Wir betrachten eine Urne mit N Objekten. Darunter sind M Objekte einer bestimmten Sorte, wir nennen sie Sorte M, die Merkmalsträger und $N - M$ andersartige Objekte. Es wird zufällig eine Stichprobe von n Objekten aus der Urne entnommen. Das Ziehen kann auf einmal passieren oder auch nacheinander, wichtig aber ist, dass *ohne Zurücklegen* der Objekte gezogen wird. Mit der Zufallsvariable X sei die *Anzahl der Merkmalsträger in der Stichprobe* bezeichnet. Die Verteilung von X heisst *hypergeometrische Verteilung*. Mit Binomialkoeffizienten lässt sich die Anzahl der relevanten Ergebnisse abzählen und die Zähldichte (PMF) der Zufallsvariablen X direkt beschreiben. Die Anzahl der Möglichkeiten genau x Merkmalsträger aus insgesamt M Merkmalsträgern zu ziehen ist

$$\binom{M}{x}$$

Eine Ziehung besteht aus n Objekten. Um insgesamt genau n Objekte zu erhalten müssen noch $n - x$ Objekte gezogen werden, welche keine Merkmalsträger sind. Dafür gibt es

$$\binom{N-M}{n-x}$$

Möglichkeiten. Insgesamt gibt es

$$\binom{M}{x} \cdot \binom{N-M}{n-x}$$

Möglichkeiten für eine Stichprobe mit x Merkmalsträgern und $n - x$ Objekten ohne Merkmal. Die Wahrscheinlichkeit für genau x Merkmalsträger in der Stichprobe von n Objekten ist demnach:

$$P(X = x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

Beispiel 1: Lotto 6 aus 49

Es soll die Wahrscheinlichkeit für genau x Treffer (x erratene Gewinnzahlen) bei einem Versuch bestimmt werden. Bei diesem Zufallsexperiment gibt es genau 6 Zahlen, welche Merkmalsträger sind, nämlich die 6 Gewinnzahlen. 43 weitere Zahlen sind keine Merkmalsträger. Wenn wir die Notation von oben benutzen, so geht es bei diesem Spiel um eine Stichprobe von $n = 6$ Zahlen aus einer Gesamtheit von $N = 49$ Zahlen. Darunter sind $M = 6$ Merkmalsträger, die Gewinnzahlen, und $N - M = 43$ Zahlen, welche das Merkmal nicht haben, die Nieten. Die Anzahl der Gewinnzahlen X in der Stichprobe ist hypergeometrisch verteilt mit der Dichte

$$P(X = x) = \frac{\binom{6}{x} \cdot \binom{43}{6-x}}{\binom{49}{6}}$$

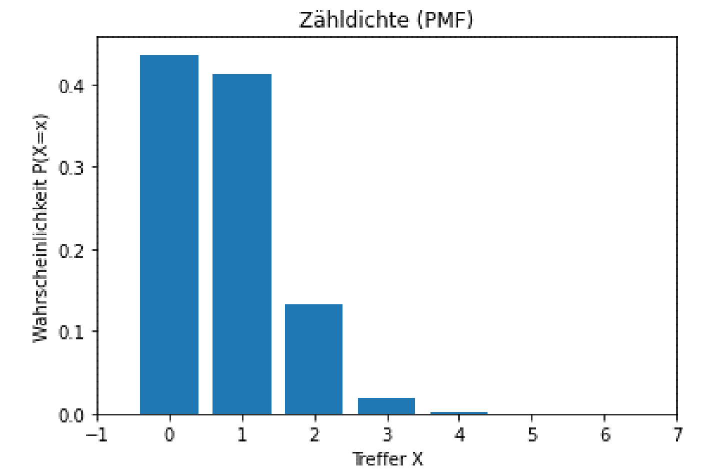
Beispielsweise ist die Wahrscheinlichkeit für mindestens 4 richtige Zahlen:

$$P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6)$$
$$= \frac{1}{\binom{49}{6}}\left[\binom{6}{4}\cdot\binom{43}{2} + \binom{6}{5}\cdot\binom{43}{1} + \binom{6}{6}\cdot\binom{43}{0}\right] \approx 9.8714 \cdot 10^{-4}.$$

Unter 1000 Rateversuchen wird ein Versuch mit mindestens 4 richtigen Zahlen erwartet. Wenn man die CDF der Verteilung zur Verfügung hat, so kann diese Wahrscheinlichkeit am besten über das Gegenereignis und die CDF $F(x) = P(X \leq x)$ der hypergeometrischen Verteilung berechnet werden.

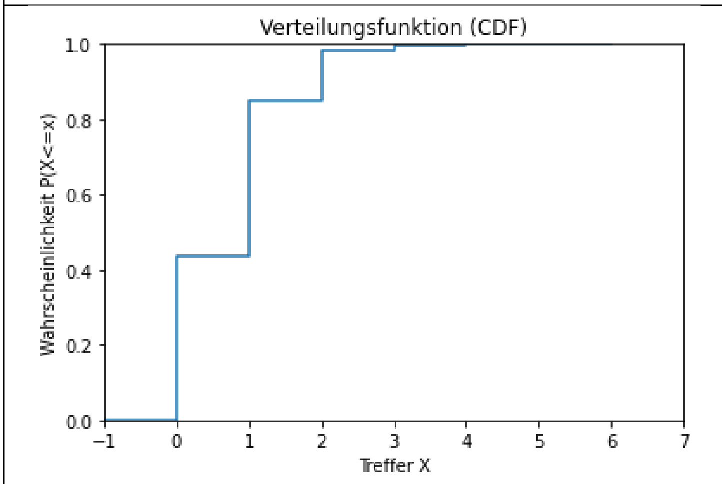
$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - F(3) \approx 9.8714 \cdot 10^{-4}$$

Im Python ScyPy-Modul wäre das der Befehl *1-ss.hypergeom(49,6,6).cdf(3)*




```
import numpy as np
import scipy.stats as ss
import matplotlib.pyplot as plt
N=49
M=6
n=6
X=np.arange(0,n+1)
hyge=ss.hypergeom(N, M, n)
f=hyge.pmf(X)
F=hyge.cdf(X)

plt.figure(1)
plt.bar(X,f)
plt.title('Zähldichte (PMF)')
plt.xlabel('Treffer X')
plt.ylabel('Wahrscheinlichkeit P(X=x)')
plt.xlim(-1,7)
```




```
plt.figure(2)
plt.step(np.append([-1],X),np.append([0],F),where='post')
plt.title('Verteilungsfunktion (CDF)')
plt.xlabel('Treffer X')
plt.ylabel('Wahrscheinlichkeit P(X<=x)')
plt.xlim(-1,7)
plt.ylim(0,1)

#Berechung von P(X≥4)
print(1-ss.hypergeom(49,6,6).cdf(3))
```

Hypergeometrische Verteilung

Eine diskrete Zufallsvariable X heisst *hypergeometrisch verteilt* mit Parametern: *n* Objekte aus einer Gesamtheit von *N* Objekten mit *M* Merkmalsträgern und *N-M* andersartigen Objekten, wenn Sie die folgende Verteilung besitzt:

$$P(X = x) = \frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}} \text{ (PMF von X)}$$

Schreibweise: $X \sim H(N, M, n)$

Die Zufallsvariable X gibt die Anzahl der Merkmalsträger in der Stichprobe an.
Python (ScyPy): *ss.hypergeom(N,M,n).pmf(x)* bzw. *ss.hypergeom(N,M,n).cdf(x)*

Erwartungswert, Varianz und Standardabweichung der Hypergeometrischen Verteilung

Für eine Zufallsvariable $X \sim H(N, M, n)$ gilt:

$\mu = E(X) = n \cdot \frac{M}{N}$, $\sigma^2 = V(X) = n \cdot \frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}$, $\sigma = S(X) = \sqrt{n \cdot \frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}}$

Beispiel 2:
In einem Behälter befinden sich 20 Kugeln, davon sind 4 blau und 16 rot. Aus dem Behälter werden nun ohne Zurücklegen 5 Kugeln zufällig entnommen.

a) Wie gross ist die Wahrscheinlichkeit, in dieser Stichprobe genau 2 blaue Kugeln vorzufinden

b) Geben Sie die Wahrscheinlichkeitsverteilung der Zufallsvariablen X = Anzahl der blauen Kugeln in der Stichprobe an. Welche Anzahl von blauen Kugeln ist einer 5-er Stichprobe am wahrscheinlichsten?

c) Geben Sie Erwartungswert und Varianz an.

a) $N = 20, M = 4, n = 5$

$$P(X = 2) = \frac{\binom{4}{2} \cdot \binom{16}{3}}{\binom{20}{5}} = 0.217$$

b) $P(X = x) = \frac{\binom{4}{x} \cdot \binom{16}{5-x}}{\binom{20}{5}}$

x	0	1	2	3	4
$P(X = x)$	0.282	0.470	0.217	0.031	0.001

Es ist also am wahrscheinlichsten, dass in der Stichprobe eine blaue Kugel gezogen wird.

c) $\mu = E(X) = 5 \cdot \frac{4}{20} = 1$

$$\sigma^2 = V(X) = 5 \cdot \frac{4}{20}\left(1 - \frac{4}{20}\right)\frac{20-5}{20-1} = 5 \cdot \frac{4}{20} \cdot \frac{16}{20} \cdot \frac{15}{19} = \frac{12}{19} \approx 0.632$$

5.2.2 Bernoulliverteilung

Bernoulli-Experimente sind Zufallsexperimente mit nur zwei möglichen Ergebnissen. Wir bezeichnen diese Ergebnisse mit 1 und 0.

Beispiele

- Qualitätskontrolle: 1 (Gerät funktioniert einwandfrei) und 0 (Gerät ist defekt).
- Abstimmung: 1 (Ja) und 0 (Nein).

Nun betrachten wir eine Bernoulli-verteilte Zufallsvariable X und verwenden folgende Bezeichnungen:

$$P(X = 1) = p \text{ und } P(X = 0) = 1 - p = q.$$

Damit können wir den Erwartungswert und die Varianz von X berechnen:

$$E(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = 1 \cdot p + 0 \cdot q = p$$

$$E(X^2) = 1^2 \cdot P(X^2 = 1) + 0^2 \cdot P(X^2 = 0) = 1 \cdot p + 0 \cdot q = p$$

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p \cdot (1 - p) = pq$$

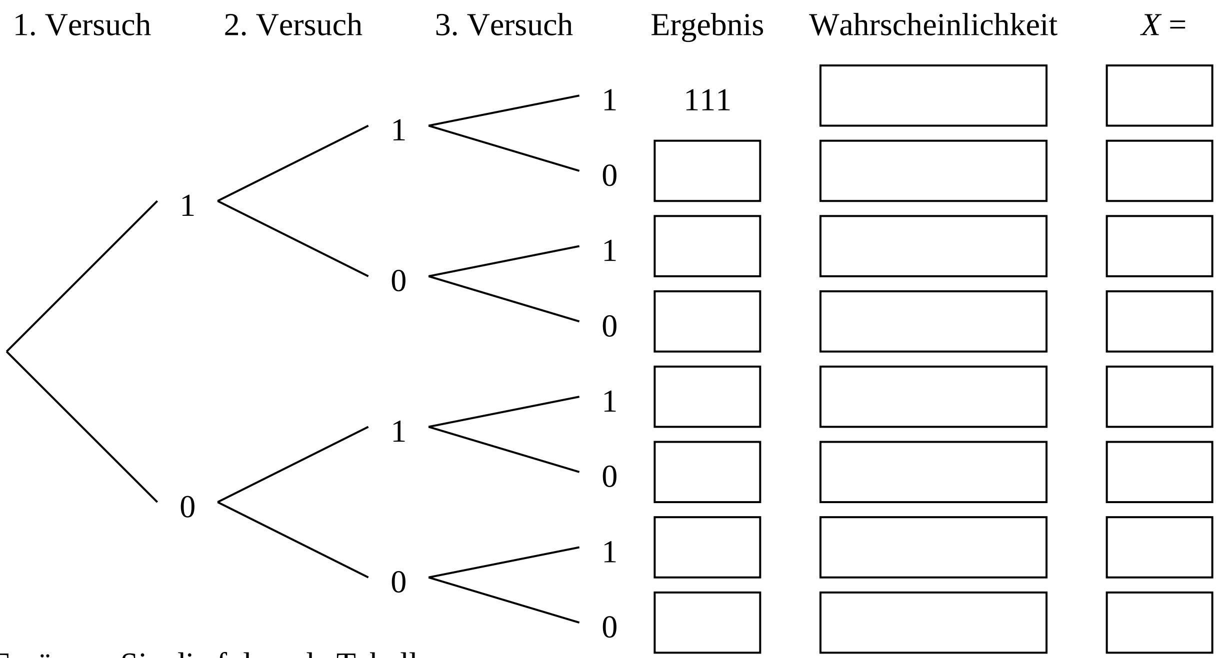
5.2.3 Binomialverteilung

Für die Binomialverteilung wird dasselbe Bernoulli-Experiment n-mal hintereinander durchgeführt; dabei sind die Wiederholungen stochastisch unabhängig. Wie gehabt bezeichnet p die Wahrscheinlichkeit für ein Ergebnis 1 und q die Wahrscheinlichkeit für ein Ergebnis 0. Die binomialverteilte Zufallsvariable X zählt, wie oft dabei das Ergebnis 1 eintritt.

Anders ausgedrückt: Wir haben n Bernoulli-verteilte Zufallsvariablen X_i, die je das Ergebnis des i-ten Experiments festhalten. Die binomialverteilte Zufallsvariable X ist dann gleich der Summe der X_i: $X = X_1 + X_2 + \ldots + X_n$.

Beispiel 3: Fall $n = 3$

Nachfolgend ist der Wahrscheinlichkeitsbaum für die 3-fache Ausführung eines Bernoulli-Experiments abgebildet. Ergänzen Sie im Wahrscheinlichkeitsbaum die Wahrscheinlichkeiten, die zu den einzelnen Verbindungslinien gehören, und füllen Sie dann die Felder auf der rechten Seite aus. Drücken Sie die Wahrscheinlichkeiten durch p und q aus (keine Zahlen einsetzen!).



Ergänzen Sie die folgende Tabelle:

x	0	1	2	3
$P(X = x)$	q^3	$3pq^2$	$3p^2q$	p^3

Allgemeiner Fall
Versuchen Sie, diese Resultate auf den allgemeinen Fall (n beliebig) zu übertragen.

1. Welche Wahrscheinlichkeit hat *ein einzelner Pfad* mit x -mal einer 1 in n Versuchen?

$$p^x \cdot q^{n-x}$$

2. Es gibt $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ verschiedene Pfade mit x -mal einer 1 in n Versuchen. Wie gross ist demnach die Wahrscheinlichkeit $P(X = x)$?

$$\binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Beispiel 4:
Die Wahrscheinlichkeit, dass ein bestimmter Gerätetyp einer Zuverlässigkeitsprüfung nicht standhält, beträgt $p = 0.06$. Es werden $n = 200$ Geräte unabhängig voneinander dieser Prüfung unterzogen. Die Anzahl X der dabei herausfallenden schlechten Geräte ist binomial-verteilt; für die Dichtefunktion PMF gilt:

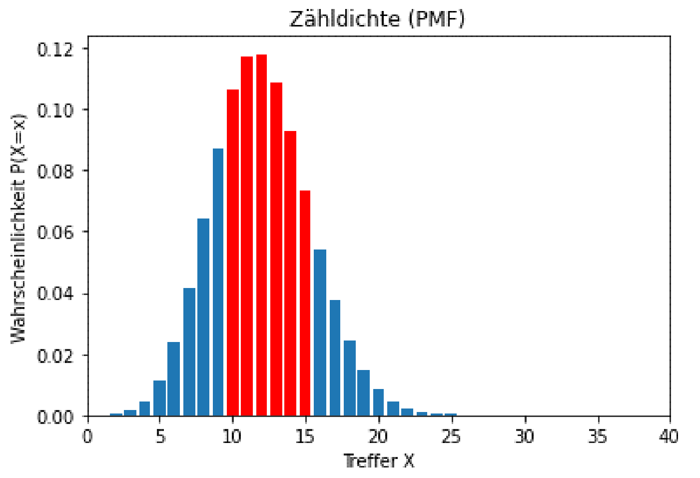
$$P(X = x) = \binom{200}{x} \cdot 0.06^x \cdot 0.94^{200-x}$$

Die Wahrscheinlichkeit, dass von diesen Geräten mindestens 10 und höchstens 15 die Zuverlässigkeitsprüfung nicht bestehen, wird mithilfe der kumulativen Verteilungsfunktion (CDF) $F(x) = P(X \leq x)$ ausgedrückt:

$$P(10 \leq X \leq 15) = P(X \leq 15) - P(X \leq 9) = F(15) - F(9)$$

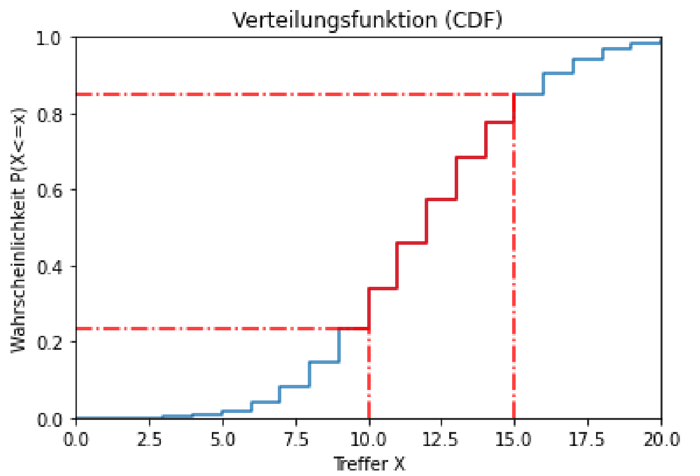
$$= \sum_{k=10}^{15} \binom{200}{k} \cdot 0.06^k \cdot 0.94^{200-k} \approx 0.6169$$

Im Python ScyPy-Modul wäre das der Befehl
ss.binom(200,0.06).cdf(15) – ss.binom(200,0.06).cdf(9)




```
import numpy as np
import scipy.stats as ss
import matplotlib.pyplot as plt
n=200
p=0.06
X=np.arange(0,n+1)
X0=np.arange(10,16)
bi=ss.binom(n,p)
f=bi.pmf(X)
F=bi.cdf(X)
f0=bi.pmf(X0)
F0=bi.cdf(X0)

plt.figure(1)
plt.bar(X,f)
plt.bar(X0,f0,color='red')
plt.title('Zähldichte (PMF)')
plt.xlabel('Treffer X')
plt.ylabel('Wahrscheinlichkeit P(X=x)')
plt.xlim(0,40)
```




```
plt.figure(2)
plt.step(np.append([-1],X),np.append([0],F),where='post')
plt.step(np.append([9],X0),np.append(F[9],F0),where='post',color='red')
plt.plot([0,10],[F[9],F[9]],'r-.')
plt.plot([0,15],[F[15],F[15]],'r-.')
plt.plot([10,10],[0,F[9]],'r-.')
plt.plot([15,15],[0,F[15]],'r-.')
plt.title('Verteilungsfunktion (CDF)')
plt.xlabel('Treffer X')
plt.ylabel('Wahrscheinlichkeit P(X<=x)')
plt.xlim(0,20)
plt.ylim(0,1)

#Berechung von P(10≤X≤15)
F1=ss.binom(200,0.06).cdf(15)
F2= ss.binom(200,0.06).cdf(9)
print(F1-F2)
```

Bemerkung
Ein Standard-Zufallsexperiment für die Binomialverteilung ist das **Ziehen mit Zurücklegen**. Wir betrachten eine Urne mit N Objekten. Darunter sind M *Merkmalsträger*, d.h. Objekte, die eine bestimmte Eigenschaft haben; die übrigen $N - M$ Objekte haben diese Eigenschaft nicht. Nun ziehen wir n-mal zufällig ein Objekt aus der Urne und legen es nach der Registrierung gleich wieder zurück. Die Wahrscheinlichkeit, dabei einen Merkmalsträger zu ziehen, ist jedes Mal $p = \frac{M}{N}$. Die Zufallsvariable X zählt, wie viele Merkmalsträger dabei gezogen werden; dann ist X binomialverteilt.

Binomialverteilung
Eine diskrete Zufallsvariable X heisst *binomialverteilt* mit Parametern n (Anzahl Wiederholungen), p (Wahrscheinlichkeit für ein Ergebnis 1) und $q = 1 - p$, wenn ihre Verteilung gegeben ist durch

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$$

Schreibweise: $X \sim B(n, p)$

X zählt, wie oft bei der n-fachen Wiederholung eines Bernoulli-Experiments das Ergebnis 1 eintritt.
Python (ScyPy): *ss.binom(n,p).pmf(x)* bzw. *ss.binom(n,p).cdf(x)*

Beispiel 5:
Es wird zwölfmal gewürfelt. Wie gross ist die Wahrscheinlichkeit für die folgenden Ereignisse:
a) Es wird genau dreimal eine Sechs geworfen.
b) Es werden höchstens drei Sechsen geworfen.
c) Es wird mindestens eine Sechs geworfen.

Die Zufallsvariable X zählt die Sechsen. Es gilt: $X \sim B(12; \frac{1}{6})$ (n,p,x farbig kennzeichnen)

a) $P(X=3) = \binom{12}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^9 = 0.1974$

b) $P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$$= \binom{12}{0} \cdot \left(\frac{1}{6}\right)^0 \cdot \left(\frac{5}{6}\right)^{12} + \binom{12}{1} \cdot \left(\frac{1}{6}\right)^1 \cdot \left(\frac{5}{6}\right)^{11} + \binom{12}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^{12} + \binom{12}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^9$$

$$= 0.1122 + 0.2692 + 0.2961 + 0.1974 = 0.8748$$

c) $P(X \geq 1) = 1 - P(X=0) = 1 - 0.1122 = 0.8878$

Den Erwartungswert und die Varianz einer $B(n,p)$-verteilten Zufallsvariable X lässt sich einfach bestimmen, wenn wir X wie oben beschrieben als Summe von Bernoulli-verteilten Zufallsvariablen X_i auffassen:

$$X = X_1 + X_2 + \ldots + X_n$$

Nach den Regeln für Kenngrössen von Zufallsvariablen erhalten wir den Erwartungswert und die Varianz von X folgendermassen:

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n \cdot p$$

$$V(X) = V\left(\sum_{i=1}^{n} X_i\right) \underbrace{=}_{X_i \text{ st. unabhängig}} \sum_{i=1}^{n} V(X_i) = n \cdot p \cdot q$$

Erwartungswert, Varianz und Standardabweichung der Binomialverteilung

Für eine Zufallsvariable $X \sim B(n,p)$ gilt:
$\mu = E(X) = np,\ \sigma^2 = V(X) = npq, \sigma = S(X) = \sqrt{npq}$

Beispiel 6:
Beim alljährlichen Blasrohr-Wettkampf wird die Anzahl der Treffer des favorisierten Mr. Q mithilfe der binomialverteilten Zufallsvariable X modelliert. Wir nehmen an, dass X den Erwartungswert 20 und die Standardabweichung 2 besitzt. Berechnen Sie die Wahrscheinlichkeit, dass Mr. Q. mehr als 23 Treffer landet.

$\boldsymbol{np = 20, npq = 4}$

$\Rightarrow \quad \boldsymbol{q = \frac{4}{20} = \frac{1}{5}, p = \frac{4}{5}, n = 20 \cdot \frac{5}{4} = 25}$

$P(X \geq 24) = \binom{25}{24}\left(\frac{4}{5}\right)^{24}\left(\frac{1}{5}\right) + \binom{25}{25}\left(\frac{4}{5}\right)^{25}\left(\frac{1}{5}\right)^{0} \approx 0.02739$

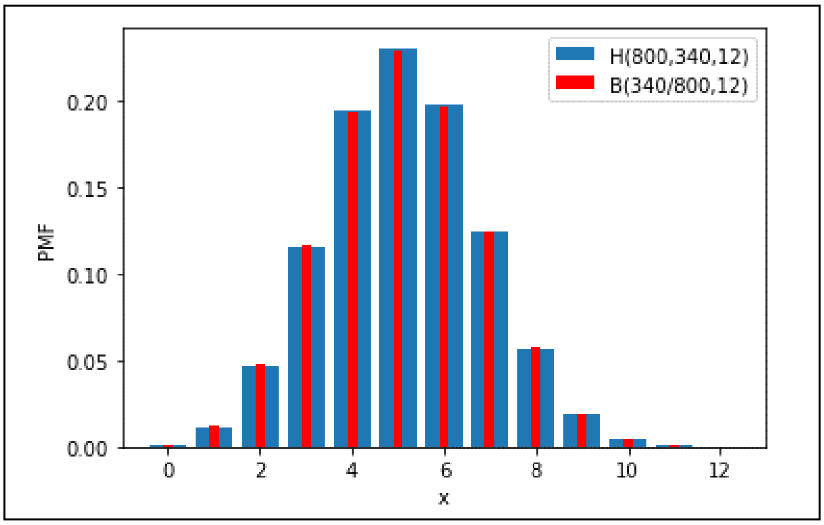
Bei n-maligem Ziehen ohne Zurücklegen aus einer endlichen, aber sehr grossen Grundgesamtheit (z. B. alle Haushalte eines Landes) ändert das Ziehen eines Elementes den Rest der Grundgesamtheit nur wenig, so dass näherungsweise angenommen werden kann, dass beim nächsten Ziehen wieder dieselben Bedingungen vorliegen.
Im diesem Fall kann eine hypergeometrisch-verteilte Zufallsvariable als nähreungsweise binomialverteilt angenommen werden.

Bemerkung:
Wenn die Bedingung (Faustregel)

$$n \lesssim \frac{N}{20}$$

erfüllt ist, dann kann die hypergeometrische Verteilung mit den Parametern n, M, N durch die rechnerisch einfachere Binomialverteilung angenähert werden: $H(N, M, n) \approx B\left(n, \frac{M}{N}\right)$

Unten im Bild ist als Beispiel die Approximation der Zähldichte (PMF) der hypergeometrischen Verteilung $H(800, 340, 12)$ durch die Binomialverteilung $B(340/800, 12)$ zu sehen.



Beispiel 7: Binomialverteilung als Näherung der hypergeometrischen Verteilung
In einer Lieferung sind 2000 Einheiten, davon sind 60 fehlerhaft. Es wird eine zufällige Stichprobe vom Umfang n = 50 entnommen. Wie gross ist die Wahrscheinlichkeit, genau zwei fehlerhafte Einheiten zu ziehen? Lösen Sie
a) exakt und
b) mit Näherung durch die Binomialverteilung.

a) Exakt: $P(X=2)=\frac{\binom{60}{2}\binom{1940}{48}}{\binom{2000}{50}}\approx 0.259$ (Wert angeben oder mit Python rechnen lassen)
b) Näherung: $P(X=2)=\binom{50}{2}(0.03)^2(1-0.03)^{48}\approx 0.256$

5.2.4 Poissonverteilung

Die *Poisson Verteilung* wird immer dort als stochastisches Modell benutzt, wo es um die Wahrscheinlichkeit für das Eintreten einer bestimmten Anzahl gleichartiger Ereignisse geht, welche in einem gegebenen Bereich (unabhängig voneinander) beliebig oft auftreten können.

Unter diesen Voraussetzungen kann zum Beispiel die Anzahl X der Unfälle pro Wochenende, die Anzahl X der Lackfehler pro cm^2 gefertigten Blechs, die Anzahl Server Anfragen X in einer Computer Zentrale, die Ausfälle X von bestimmten Bauteilen oder die Meteoriteneinschläge X in einem bestimmten räumlichen Bereich oder einer zeitlichen Periode poissonverteilt sein.

Ein weiteres Beispiel werden wir nun verwenden, um die Verteilung genauer zu beschreiben: die Anrufzahlen in einer Telefonzentrale.
Wir betrachten eine Telefonzentrale in einem Zeitintervall $[0, T]$. Die Zufallsvariable X gebe die Anzahl der in diesem Zeitintervall erfolgten Anrufe an. Wir wollen eine Verteilung für diese Zufallsvariable herleiten. Dazu machen wir die folgenden Annahmen:

1. Anrufe sind im gegebenen Zeitintervall unabhängig voneinander;
2. Anrufe sind im gegebenen Zeitintervall gleichverteilt (es gibt im Zeitintervall keine Stoss- oder Niederzeiten);
3. Die durchschnittliche Zahl der Anrufe $\lambda > 0$ im gegebenen Zeitintervall ist bekannt.

Dann wählen wir eine natürliche Zahl $n > 0$ und teilen das Zeitintervall in n gleichgrosse Teilintervalle der Länge T/n. Die Zahl n sei so gross, dass in jedem Teilintervall immer nur höchstens ein Anruf erfolgen kann.
Da es durchschnittlich $\lambda > 0$ Anrufe im Zeitintervall gibt und die Anrufe gleichverteilt sind, nehmen wir weiter an, dass in jedem Teilintervall ein Anruf mit der Wahrscheinlichkeit $p = \lambda/n$ erfolgt. Wegen der Unabhängigkeit der Anrufe, ist die Anzahl X der Anrufe dann binomialverteilt mit der Zähldichte (PMF)
$P(X=x)=\binom{n}{x}\cdot\left[\frac{\lambda}{n}\right]^x\cdot\left[1-\frac{\lambda}{n}\right]^{n-x}$.

Der Nachteil einer solchen Binomialverteilung ist, dass nicht mehr als $n > 0$ Anrufe im gegebenen Zeitintervall erfolgen können. Eine solche obere Schranke wollen wir nicht festlegen. Aus diesem Grund werden wir (bei konstantem $x \geq 0$) in einem Grenzprozess den Parameter $n > 0$ immer grösser werden lassen, d.h. das gegebene Zeitfenster $[0, T]$ in immer kleinere Teilintervalle zerlegen. Die resultierende Grenzverteilung dieser Binomialverteilungen wird, unter den getroffenen Annahmen, jede beliebige Anrufzahl im gegebenen Zeitintervall zulassen:

$$\binom{n}{x} \cdot [\frac{\lambda}{n}]^x \cdot [1 - \frac{\lambda}{n}]^{n-x} = \frac{n! \cdot \lambda^x \cdot [1 - \frac{\lambda}{n}]^n}{(n-x)! \cdot x! \cdot n^x \cdot [1 - \frac{\lambda}{n}]^x}$$

$$= \underbrace{\frac{n(n-1)\ldots(n-x+1)}{n \cdot n \cdot \ldots \cdot n}}_{\to 1} \cdot \underbrace{\frac{1}{[1 - \frac{\lambda}{n}]^x}}_{\to 1} \cdot \underbrace{[1 - \frac{\lambda}{n}]^n}_{\to e^{-\lambda}} \cdot \frac{\lambda^x}{x!}$$

Im Grenzprozess $n \to \infty$ erhält man eine neue Zähldichte (PMF) :

$$P(X = x) = \binom{n}{x} \cdot [\frac{\lambda}{n}]^x \cdot [1 - \frac{\lambda}{n}]^{n-x} \xrightarrow{n \to \infty} e^{-\lambda} \cdot \frac{\lambda^x}{x!}.$$

Also setzen wird für die Wahrscheinlichkeit für genau x Anrufe im Zeitintervall $[0, T]$:

$$P(X = x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}.$$

Dies ist die Zähldichte (PMF) der sogenannten *Poissonverteilung*. Sie ist, wie oben dargestellt, als Grenzverteilung von Binomialverteilungen definiert.

> **Poissonverteilung**
> Eine diskrete Zufallsvariable X heisst *poissonverteilt* mit dem Parametern λ, wenn ihre Verteilung gegeben ist durch
>
> $P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \lambda > 0$
>
> Schreibweise: $X \sim Poi(\lambda)$

Die Zufallsvariable X beschreibt die Anzahl der Ereignisse je Zeit, Fläche, Länge, etc.
Python (ScyPy): *ss.poisson(λ).pmf(x)* bzw. *ss.poisson(λ).cdf(x)*

Beispiel 8:

Angenommen, die Anzahl X der zur Hauptgeschäftszeit eintretenden Anrufe in einer Telefonzentrale ist poissonverteilt. Die Zentrale erhält (im Mittel) 120 Anrufe pro Stunde. Wie gross ist die Wahrscheinlichkeit, dass innerhalb einer Minute

a) kein Anruf
b) genau ein
c) höchstens drei
d) mehr als drei Anrufe eintreffen?

Es gilt also: $\lambda = 2 \frac{1}{min}$

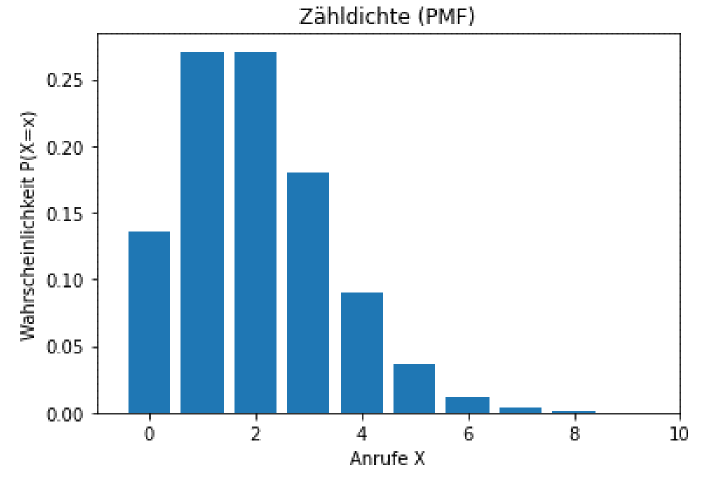
a) $P(X = 0) = \frac{2^0}{0!} e^{-2} = e^{-2} \approx 0.135$

b) $P(X = 1) = \frac{2^1}{1!} e^{-2} = 2e^{-2} \approx 0.271$

c) $P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} + \frac{2^2}{2!}e^{-2} + \frac{2^3}{3!}e^{-2} = \left(1+2+2+\frac{4}{3}\right)e^{-2} \approx 0.857$

d) $P(X > 3) = 1 - P(X \leq 3) \approx 0.143$

Graphisch dargestellt ergibt sich:




```
import numpy as np
import scipy.stats as ss
import matplotlib.pyplot as plt

l=2
n=10
X=np.arange(0,n+1)
f=ss.poisson(l).pmf(X)
F=ss.poisson(l).cdf(X)

plt.figure(1)
plt.bar(X,f)
plt.title('Zähldichte (PMF)')
plt.xlabel('Anrufe X')
plt.ylabel('Wahrscheinlichkeit P(X=x)')
plt.xlim(-1,10)
```

Die Poisson-Verteilung ist eingipfelig (unimodal). Je kleiner λ ist, desto weiter links liegt der Modalwert. Für grösseres λ, etwa ab $\lambda \geq 10$, wird die Verteilung annähernd symmetrisch.

Erwartungswert, Varianz und Standardabweichung der Poissonverteilung

Für eine Zufallsvariable $X \sim Poi(\lambda)$ gilt:
$\mu = E(X) = \lambda,\ \sigma^2 = V(X) = \lambda, \sigma = S(X) = \sqrt{\lambda}$

Beispiel 9:
Bei der Herstellung von optischen Speichermedien treten störende Staubteilchen auf. Es wird angenommen, dass die Anzahl der Staubteilchen poissonverteilt ist, mit durchschnittlich 0.05 Staubteilchen pro cm^2. Wie grosss ist die Wahrscheinlichkeit, auf einer CD von 100 cm^2 weniger als 3 Staubteilchen zu finden?
Wie gross sind der Erwartungswert und die Varianz?

Es gilt also: $\lambda = 5 \frac{1}{100\ cm^2}$

$$P(X<3) = P(X=0) + P(X=1) + P(X=2) = \frac{5^0}{0!}e^{-5} + \frac{5^1}{1!}e^{-5} + \frac{5^2}{2!}e^{-5} = \left(1+5+\frac{25}{2}\right)e^{-5} \approx 0.125$$

$$E(X) = Var(X) = 5\frac{1}{100\ cm^4}$$

Die Poisson-Verteilung wird in der Praxis oft als Näherung der Binomialverteilung verwendet.

$$B(n,p) \xrightarrow{n\to\infty \text{ und } \lambda=np \text{ konstant}} Poi(\lambda)$$

Für grosses n und kleines p kann demnach die Binomialverteilung durch die Poisson Verteilung approximiert werden:

$$B(n,p) \approx Poi(n\cdot p)$$

Faustregel: $n \gtrsim 50$ und $p \lesssim 0.1$

Beispiel 10:
Eine Fabrik produziert Werkstücke mit einem konstanten Ausschussanteil von $p = 0.002$. Wie gross ist die Wahrscheinlichkeit, dass eine Lieferung von $n = 800$ Werkstücken höchstens zwei defekte Werkstücke enthält? Berechnen Sie diese Wahrscheinlichkeit
a) exakt
b) näherungsweise mithilfe der Poissonverteilung.

a) Zufallsvariable ist also binomialverteilt mit $n = 800, p = 0.002$

$$P(X \le 2) = P(X=0) + P(X=1) + P(X=2) = \sum_{k=0}^{2} \binom{800}{k} \cdot 0.002^k \cdot 0.998^{800-k} \approx 0.78336$$

b) Die Zufallsvariable X kann als annähernd poissonverteilt betrachtet werden (da $n = 800 > 50$ und $p = 0:002 < 0:1$ ist), mit dem Parameter $\lambda = 800 \cdot 0.02 = 1.6$.

$$P(X \le 2) = P(X=0) + P(X=1) + P(X=2) = \frac{1.6^0}{0!}e^{-1.6} + \frac{1.6^1}{1!}e^{-1.6} + \frac{1.6^2}{2!}e^{-1.6}$$
$$= (1 + 1.6 + 0.64)e^{-1.6} \approx 0.78346$$

5.2.5 Zusammenfassung

Für die drei wichtigsten diskreten Verteilungen sind in der folgenden Tabelle die Dichtefunktion PMF sowie der Erwartungswert $E(X)$, die Varianz $V(X)$ und die Standardabweichung $S(X)$, in Abhängigkeit der jeweiligen Parameter zu sehen.

	Hypergeometrische Vert.	Binomialverteilung	Poissonverteilung
Dichtefunktion für $x = 0,1,2,\dots$ sonst 0	$P(X=x) = \dfrac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}}$	$P(X=x) = \binom{n}{x}\cdot p^x\cdot(1-p)^{n-x}$	$P(X=x) = \dfrac{\lambda^x}{x!}e^{-\lambda}$
Erwartungs-wert	$n\cdot\frac{M}{N}$	$n\cdot p$	λ
Varianz	$n\cdot\frac{M}{N}\cdot(1-\frac{M}{N})\cdot\frac{N-n}{N-1}$	$n\cdot p\cdot(1-p)$	λ
Standardab-weichung	$\sqrt{n\cdot\frac{M}{N}\cdot(1-\frac{M}{N})\cdot\frac{N-n}{N-1}}$	$\sqrt{n\cdot p\cdot(1-p)}$	$\sqrt{\lambda}$
Parameter	n, N, M	n, p	λ

5.3 Stetige Verteilungen

5.3.1 Gauss- oder Normalverteilung

Die Normalverteilung ist die bedeutendste Verteilung, da viele natürliche Prozesse mit ihr modelliert werden können. Der Zentrale Grenzwertsatz gibt darüber Aufschluss. Im Wesentlichen besagt dieses von Carl Friedrich Gauss entdeckte Theorem, dass das arithmetische Mittel identisch verteilter stochastisch unabhängiger Zufallsvariablen annähernd normalverteilt ist und mit wachsender Anzahl der Zufallsvariablen immer mehr mit der Normalverteilung übereinstimmt. Gauss hat damit die zentrale Bedeutung der Normalverteilung inmitten aller anderen Verteilungen entdeckt.

Definition
Die stetige Zufallsvariable X folgt der *Normalverteilung* mit den Parametern $\mu, \sigma \in \mathbb{R}, \sigma > 0$, wenn sie folgende Dichtefunktion hat:

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

Schreibweise: $X \sim N(\mu; \sigma)$

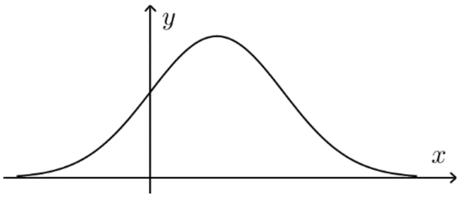
Ist $\mu = 0$ und $\sigma = 1$, so spricht man von der *Standardnormalverteilung*. Ihre Dichtefunktion wird einfach mit φ bezeichnet; sie ist gegeben durch:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$$

Schreibweise: $X \sim N(0; 1)$

Python (ScyPy): *ss.norm(λ).pdf(x)* bzw. *ss.norm(λ).cdf(x)*

Bemerkung
Die Dichtefunktion $\varphi_{\mu,\sigma}(x)$ hat folgende Eigenschaften:



(a) Sie ist symmetrisch bezüglich der Geraden $x = \mu$.
(b) Sie hat Wendepunkte an den Stellen $\mu - \sigma$ und $\mu + \sigma$.
(c) Sie ist *normiert*, d.h. es gilt:Eigenschaften einzeichnen

$$\int_{-\infty}^{\infty} \varphi_{\mu,\sigma}(x)\, dx = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

(d) Eine Änderung von μ bewirkt eine Verschiebung in x-Richtung; je grösser σ ist, desto breiter und niedriger wird die Glockenkurve.

Erwartungswert und Varianz der Normalverteilung
Für eine Zufallsvariable $X \sim N(\mu; \sigma)$ gilt

$$E(X) = \mu,\ V(X) = \sigma^2$$

Die Verteilungsfunktion der Normalverteilung
Die kumulative Verteilungsfunktion (CDF) von $\varphi_{\mu,\sigma}(x)$ wird mit $\phi_{\mu,\sigma}(x)$ bezeichnet. Sie ist definiert durch:

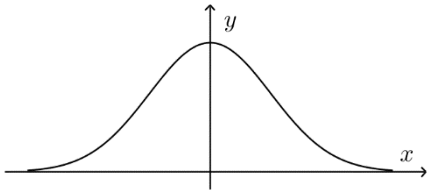
$$\phi_{\mu,\sigma}(x) = P(X \leq x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma}(t)\,dt = \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}\,dt$$

Die Verteilungsfunktion der Normalverteilung kann nicht auf elementare Weise berechnet werden. Für ihre Werte gibt es Tabellen (Papula, 12. Aufl., S. 514); die Tabellen beziehen sich allerdings immer auf die *Standard*normalverteilung.

Beispiel 11:
Gegeben ist eine standardnormalverteilte Zufallsvariable X. Bestimmen Sie

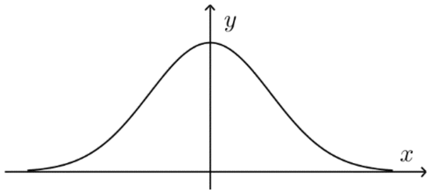
(a) $P(X \leq 0.45) =$ 0.6736

(b) $P(X \leq 2.37) =$ 0.9911



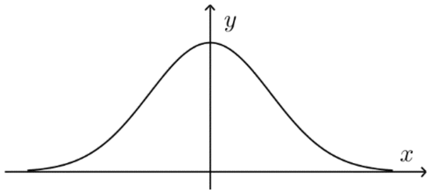
(c) $P(X \leq -0.21) =$ $1-\phi(0.21)$

$=1-0.5832=0.4168$



(d) $P(0.17 \leq X \leq 1.85) =$ $\phi(1.85)-\phi(0.17)$

$=0.9678-0.5675=0.4003$

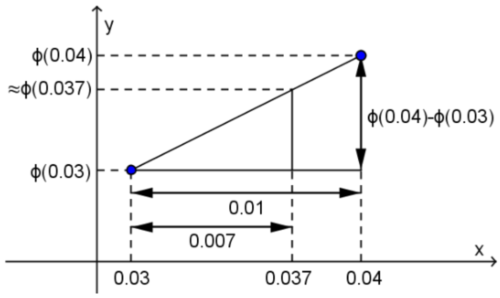


Hat die vorgegebene Grenze mehr als zwei Nachkommastellen, wird linear interpoliert:
(e) $P(X \leq 0.037)$

$=\phi(0.03)+0.007\cdot\dfrac{\phi(0.04)-\phi(0.03)}{0.01}$

$=0.5120+0.7\cdot(0.5160-0.5120)$

$=0.5120+0.0028=0.5148$



Liegt eine beliebige Normalverteilung $N(\mu;\sigma)$ vor, muss erst *standardisiert* werden: Statt der ursprünglichen Zufallsvariablen X betrachtet man die Zufallsvariable

$$U = \frac{X-\mu}{\sigma}$$

Diese folgt der Standardnormalverteilung.

Beispiel 12:
Die Zufallsvariable X folgt der Verteilung $N(\mu;\sigma)$ mit $\mu = 3$ und $\sigma = 2$. Bestimmen Sie

(a) $P(X \leq 4.84) = P\left(\frac{X-\mu}{\sigma} \leq \frac{4.84-\mu}{\sigma}\right) = P\left(U \leq \frac{4.84-3}{2}\right) = \phi(0.92) = 0.8212$

(b) $P(1.26 \leq X \leq 5.12) = P\left(\frac{1.26-3}{2} \leq U \leq \frac{5.12-3}{2}\right) = \phi(1.06) - (1-\phi(0.87))$

$= 0.8554 - 1 + 0.8078 = 0.6632$

Häufig braucht man die Wahrscheinlichkeit dafür, dass eine normalverteilte Zufallsvariable Werte in dem symmetrisch zu μ gelegenen Intervall $[\mu-e;\mu+e]$ mit $e>0$ annimmt. Es gilt:

$$P(\mu-e \leq X \leq \mu+e) = P\left(-\frac{e}{\sigma} \leq U \leq \frac{e}{\sigma}\right) = \phi\left(\frac{e}{\sigma}\right) - \left(1-\phi\left(\frac{e}{\sigma}\right)\right) = 2\cdot\phi\left(\frac{e}{\sigma}\right) - 1$$

Es ist üblich, die Abweichung e von μ in ganzzahligen Vielfachen von σ anzugeben. Speziell für $e = k\sigma$ mit $k = 1,2,3$, ergeben sich folgende Werte:

$P(\mu-\sigma \leq X \leq \mu+\sigma) = 2\phi(1) - 1 = 0.6826$

$P(\mu-2\sigma \leq X \leq \mu+2\sigma) = 2\phi(2) - 1 = 0.9544$

$P(\mu-3\sigma \leq X \leq \mu+3\sigma) = 2\phi(3) - 1 = 0.9974$

Hinweis:
Werte aus der Tabelle genutzt; wenn man mit Python arbeitet, ändert sich die 4. Nachkommastelle

Satz
Bei einer Zufallsvariable X, die der Normalverteilung $N(\mu;\sigma)$ folgt, liegen
- ca. 68 % der beobachteten Werte zwischen $\mu-\sigma$ und $\mu+\sigma$,
- ca. 95 % der beobachteten Werte zwischen $\mu-2\sigma$ und $\mu+2\sigma$,
- ca. 99.7 % der beobachteten Werte zwischen $\mu-3\sigma$ und $\mu+3\sigma$.

5.3.2 Zentraler Grenzwertsatz

Für eine Folge $X_1, X_2, \ldots, X_n, \ldots$ von Zufallsvariablen definieren wir die n-te *Summe* $S_n = X_1 + \ldots + X_n$ und das *arithmetische Mittel* $\bar{X}_n = \frac{1}{n}[X_1 + \ldots + X_n]$.

Für diese beiden neuen Zufallsvariablen gilt wegen der Linearität des Erwartungswertes:
$E(S_n) = E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n)$ und
$E(\bar{X}_n) = E(\frac{1}{n}[X_1 + \ldots + X_n]) = \frac{1}{n}[E(X_1) + \ldots + E(X_n)]$

Sind die Zufallsvariablen paarweise stochastisch unabhängig kann Ähnliches für die Varianz geschlossen werden:
$V(S_n) = V(X_1 + \ldots + X_n) = V(X_1) + \ldots + V(X_n)$ und
$V(\bar{X}_n) = V(\frac{1}{n}[X_1 + \ldots + X_n]) = \frac{1}{n^2}[V(X_1) + \ldots + V(X_n)]$.

Haben zudem alle Zufallsvariablen denselben Erwartungswert $E(X_i) = \mu$ und dieselbe Varianz $V(X_i) = \sigma^2$ so folgt:

$$E(S_n) = n \cdot \mu, \qquad V(S_n) = n \cdot \sigma^2, \qquad E(\bar{X}_n) = \mu, \qquad V(\bar{X}_n) = \frac{\sigma^2}{n}$$

Sind die Zufallsvariablen alle identisch $N(\mu, \sigma)$ -verteilt, so ist die Summe $S_n = X_1 + \ldots + X_n$ wieder normalverteilt, nämlich $N(n \cdot \mu, \sqrt{n} \cdot \sigma)$-verteilt.

Das arithmetische Mittel $\bar{X}_n = \frac{1}{n}[X_1 + \ldots + X_n]$ ist $N(\mu, \sigma/\sqrt{n})$-verteilt.

Zentraler Grenzwertsatz

Gegeben sind lauter identisch verteilte und stochastisch unabhängige Zufallsvariablen $X_1, X_2, \ldots$, alle mit demselben Erwartungswert μ und derselben Varianz σ^2. Dann hat die Summe

$$S_n = \sum_{i=1}^{n} X_i$$

den Erwartungswert $n\mu$ und die Varianz $n\sigma^2$. Die Verteilungsfunktion $F_n(u)$ der dazugehörigen standardisierten Zufallsvariable

$$U_n = \frac{(X_1 + X_2 + \ldots + X_n) - n\mu}{\sqrt{n} \cdot \sigma} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

konvergiert für $n \to \infty$ gegen die Verteilungsfunktion $\phi(u)$ der Standardnormalverteilung:

$$\lim_{n \to \infty} F_n(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{u} e^{-\frac{1}{2}t^2}\, dt$$

Eine Simulation zum Zentralen Grenzwertsatz finden Sie unter: http://onlinestatbook.com/stat_sim/sampling_dist/

In der Regel formuliert man den zentralen Grenzwertsatz, so wie hier, für die standardisierte Summe U_n der X_i. Damit sind aber auch die Summe $S_n = \sum_{i=1}^{n} X_i$ und das arithmetische Mittel $\bar{X}_n$ für $n \to \infty$ immer annähernder normalverteilt, denn alle drei Zufallsvariablen unterscheiden sich nur um eine lineare Transformation.
Die Approximation von U_n (bzw. S_n und $\bar{X}_n$) wird durch eine Normalverteilung mit entsprechenden Parametern $\mu_{(n)}, \sigma_{(n)}$ für $n \to \infty$ immer besser.

Bemerkung

Wir haben gesehen, dass das arithmetische Mittel $\bar{X}_n$ von n identisch verteilten, unabhängigen (i.i.d.) Zufallsvariablen näherungsweise $N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$-verteilt ist.

Das bedeutet: $\bar{X}_n$ hat denselben Erwartungswert, aber eine kleinere Standardabweichung als X_i. Deswegen ist es günstig, dieselbe Messung mehrfach durchzuführen und dann den Mittelwert der einzelnen Messwerte zu nehmen.
Nehmen wir an, wir erheben 1000 Messreihen à n Messungen und berechnen für jede Messreihe den Mittelwert. Wenn wir ein Histogramm dieser Mittelwerte erstellen, so wird das Histogramm umso schmaler sein, je grösser n ist. Man kann also erwarten, dass der Mittelwert der Messreihe umso näher beim wahren Wert der Messgrösse liegt, je grösser n ist (wenn ein Messfehler rein zufallsbedingt ist, also kein systematischer Messfehler vorliegt).

Bemerkung
Wesentlich an den Voraussetzungen zum zentralen Grenzwertsatz ist, dass die X_i nicht normalverteilt zu sein brauchen, sondern nur unabhängig und identisch verteilt sein müssen.
Es gibt Verallgemeinerungen dieses Satzes, die nicht verlangen, dass die Zufallsvariablen identisch verteilt sind. Damit gilt also salopp gesagt: *Eine Grösse ist näherungsweise normalverteilt, wenn sie von einer Überlagerung von vielen unabhängigen zufälligen Einflüssen abhängt* (z.B. Messgrösse). Das ist die Erklärung dafür, dass die Normalverteilung so verbreitet ist.

Beispiel 13:
Die Fertigung eines Artikels besteht aus 6 Arbeitsgängen. T_i bezeichnet die Dauer des i-ten Arbeitsganges in Stunden. Jedes T_i ist eine Zufallsvariable, die im Intervall [1;2] gleichverteilt ist. Die Unabhängigkeitsbedingung für die T_i ist erfüllt.
Bestimmen Sie näherungsweise die Verteilung der Gesamtproduktionsdauer $T = T_1 + T_2 + \ldots + T_6$ und berechnen Sie damit die Wahrscheinlichkeit, dass $T > 10$ ist.

T_i hat die Dichte 1 für $1 < t < 2$ und 0 sonst. Wegen der Symmetrie gilt: $\mu = 1.5$. Für die Varianz gilt:

$$\sigma^2 = \int_1^2 \left(t - \frac{3}{2}\right)^2 \cdot 1\, dt = \left[\frac{t^3}{3} - \frac{3}{2}t^2 + \frac{9}{4}t\right]_1^2 = \frac{7}{3} - \frac{9}{2} + \frac{9}{4} = \frac{1}{12}$$

T hat daher den Erwartungwert $6 \cdot 1.5 = 9$ und die Varianz $6 \cdot \frac{1}{12} = \frac{1}{2}$.

Nach dem Zentralen Grenzwertsatz ist T annähernd normalverteilt.

$$P(T > 10) = 1 - P(T \leq 10) = 1 - \phi\left(\frac{10-9}{\sqrt{\frac{1}{2}}}\right) = 1 - \phi(1.414) = 1 - 0.9213 = 0.0787$$

(interpoliert!)

Approximation der Binomialverteilung durch die Normalverteilung
Wichtiger Spezialfall des Zentralen Grenzwertsatzes.
Gegeben ist eine Zufallsvariable $X \sim B(n;p)$. X zählt die Ergebnisse 1 bei der n-fachen Wiederholung eines Bernoulli-Experiments. Wir haben bereits gesehen, dass man X als Summe von n stochastisch unabhängigen, Bernoulli-verteilten Zufallsvariablen X_i auffassen kann: $X = X_1 + X_2 + \ldots + X_n$. Dabei haben die Zufallsvariablen X_i alle Erwartungswert p und Varianz pq.

Damit kann man auf X den Zentralen Grenzwertsatz anwenden. Danach gilt:
X ist näherungsweise normalverteilt mit $\mu = np$ und $\sigma^2 = npq$:

$$P(a \leq X \leq b) = \sum_{x=a}^{b} P(X = x) \approx \phi_{\mu,\sigma}(b + \frac{1}{2}) - \phi_{\mu,\sigma}(a - \frac{1}{2})$$

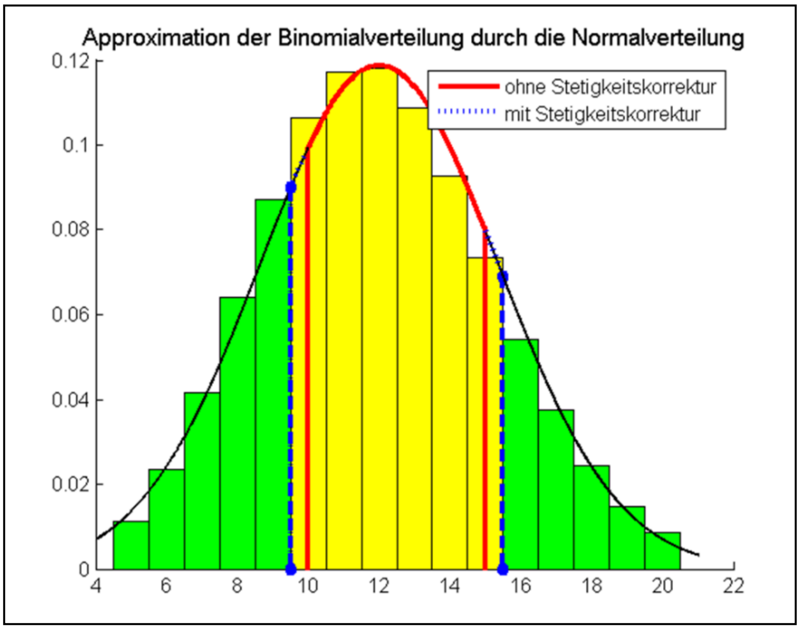
Faustregel: Diese Approximation kann verwendet werden, wenn $npq > 9$.

An den Intervallgrenzen a und b ist hier die sogenannte *Stetigkeitskorrektur* vorgenommen worden. Dies erklären wir am nächsten Beispiel .

Beispiel 14:
Die Wahrscheinlichkeit dafür, dass ein bestimmter Gerätetyp einer Zuverlässigkeitsprüfung nicht standhält, beträgt $p = 0.06$. Es werden insgesamt 200 Geräte unabhängig voneinander dieser Prüfung unterzogen. Berechnen Sie die Wahrscheinlichkeit, dass von diesen Geräten mindestens 10 und höchstens 15 diese Zuverlässigkeitsprüfung nicht bestehen.

Betrachtet man die beiden Verteilungen im Bild unten, so wird die Problematik deutlich, die entsteht, wenn allgemein eine diskrete Verteilung durch eine stetige Verteilung approximiert wird. Die gelb gefärbte Fläche entspricht der gesuchten Wahrscheinlichkeit $P(10 \leq X \leq 15)$ und der rot umrandete Bereich der approximierenden Fläche unter der Gauss'schen Glockenkurve. Die Approximation wird dabei erheblich verbessert, wenn die *Stetigkeits-korrektur* links und rechts am Integrationsbereich vorgenommen wird: Statt von 10 bis 15, wird von 9.5 bis 15.5 integriert.
Dies entspricht dem Inhalt der blau gestrichelt eingegrenzten Fläche.

$$P(10 \leq X \leq 15) = \sum_{x=a}^{b} P(X = x) \approx \phi_{12,\sqrt{11.28}}(15.5) - \phi_{12,\sqrt{11.28}}(9.5) \approx 0.6230$$



Was tun, wenn allgemein $P(a < X \leq b)$ gesucht ist?
Wir berechnen die Fläche zwischen $a + ½$ und $b + ½$.
…………………………………………………………………………………………………

Beispiel 15
Bei einer digitalen Übertragung ist die Wahrscheinlichkeit, dass ein Bit fehlerhaft übertragen wird, immer gleich 10^{-5}. Wie gross ist die Wahrscheinlichkeit, dass unter 10 Millionen Bit mehr als 90 Bit fehlerhaft übertragen werden?

$\sigma^2 = npq = 99.999 > 9$, die Bedingung der Faustregel ist also erfüllt und wir dürfen die Normalapproximation verwenden
Es gilt: $\mu = np = 100;\ \sigma^2 = npq = 99.999$

$$P(X > 90) = 1 - P(X \leq 90) = 1 - \phi_{\mu,\sigma}(90.5) = 1 - \phi\left(\frac{90.5 - 100}{\sqrt{99.999}}\right)$$
$$= 1 - \phi(-0.95) = \phi(0.95) \approx 0.829$$

Approximation der Poissonverteilung durch die Normalverteilung
Sei X poissonverteilt mit dem Parameter λ. Wenn λ gross genug ist, gilt:
X ist näherungsweise normalverteilt mit $\mu = \lambda$ und $\sigma^2 = \lambda$:

$$P(a \leq X \leq b) = \sum_{x=a}^{b} P(X = x) \approx \phi_{\mu,\sigma}(b + \frac{1}{2}) - \phi_{\mu,\sigma}(a - \frac{1}{2})$$

Auch hier verbessert die Stetigkeitskorrektur die Güte der Approximation an den Sprungstellen der Poissonverteilung

Beispiel 16
Die Anzahl von Staubteilchen pro Volumeneinheit sei poissonverteilt mit dem Erwartungswert 20. Wie gross ist die Wahrscheinlichkeit, weniger als 15 Staubteilchen pro Volumeneinheit vorzufinden?

$\sigma^2 = \lambda = 20 > 9$, die Bedingung der Faustregel ist also erfüllt und wir dürfen die Normalapproximation verwenden.
Es gilt: $\mu = 20;\ \sigma^2 = \lambda = 20$

$$P(X \leq 14) = P(X \leq 14) = \phi_{\mu,\sigma}(14.5) = \phi\left(\frac{14.5 - 20}{\sqrt{20}}\right) = \phi(-1.23) = 1 - \phi(1.23)$$
$$\approx 0.109$$

zum Vergleich: Poisson liefert 0.10486

5.4 Lernziele für dieses Kapitel

- ☐ Ich kann beurteilen, ob eine vorgegebene Zufallsvariable diskret oder stetig ist.
- ☐ Ich kann mithilfe der CDF Intervallwahrscheinlichkeiten für stetige Zufallsvariablen berechnen.
- ☐ Ich kenne die Formeln für den Erwartungswert und die Varianz von stetigen Zufallsvariablen.
- ☐ Ich kann die Unterschiede zwischen diskreten und stetigen Zufallsvariablen bezüglich der Definition und der graphischen Darstellung der PMF bzw. PDF sowie bezüglich der Berechnung von Wahrscheinlichkeiten, Erwartungswert und Varianz beschreiben.
- ☐ Ich kann den Erwartungswert und die Varianz von Bernoulli-verteilten Zufallsvariablen berechnen.
- ☐ Ich kann beurteilen, ob eine Zufallsvariable hypergeometrisch, binomialverteilt oder poissonverteilt ist.
- ☐ Ich kann erklären, wie eine binomialverteilte Zufallsvariable als Summe von Bernoulli-verteilten Zufallsvariablen definiert werden kann.
- ☐ Ich kann für hypergeometrischverteilte Zufallsvariablen Wahrscheinlichkeiten mithilfe der PMF und der CDF als Formel aufschreiben und berechnen.
- ☐ Ich kann den Erwartungswert, die Varianz und die Standardabweichung für hypergeometrischverteilte Zufallsvariablen berechnen.
- ☐ Ich kann für binomialverteilte Zufallsvariablen Wahrscheinlichkeiten mithilfe der PMF und der CDF als Formel aufschreiben und berechnen.
- ☐ Ich kann den Erwartungswert, die Varianz und die Standardabweichung für binomialverteilte Zufallsvariablen berechnen.
- ☐ Ich kann für poissonverteilte Zufallsvariablen Wahrscheinlichkeiten mithilfe der PMF und der CDF als Formel aufschreiben und berechnen.
- ☐ Ich kann den Erwartungswert, die Varianz und die Standardabweichung für poissonverteilte Zufallsvariablen berechnen.
- ☐ Ich kann die Gleichung der PDF der Normalverteilung aufschreiben.
- ☐ Ich kenne die Bedeutung der Parameter bei der Normalverteilung und kann diese am Graphen der PDF erklären.
- ☐ Ich kann erklären, was die Standardisierung einer normalverteilten Zufallsvariable bewirkt.
- ☐ Ich kann mit der Quantilstabelle der Standardnormalverteilung Intervallwahrscheinlichkeiten für beliebige normalverteilte Zufallsvariablen berechnen.
- ☐ Ich kann die Bedeutung des Zentralen Grenzwertsatzes erklären.

- ☐ Ich kenne die Verteilungen der Summe und des arithmetischen Mittels von Zufallsvariablen bei der Approximation mit dem Zentralen Grenzwertsatz.
- ☐ Ich kann den Zentralen Grenzwertsatz zur näherungsweisen Lösung von Aufgaben anwenden.
- ☐ Ich kann den Zentralen Grenzwertsatz einsetzen, um binomialverteilte Zufallsvariablen mit der Normalverteilung zu approximieren.
- ☐ Ich kann beurteilen, wann die Faustregel für die Approximation der Binomialverteilung erfüllt ist.
- ☐ Ich kann bei der Approximation der Binomialverteilung eine Stetigkeitskorrektur durchführen und deren Bedeutung am Graphen erklären.